Étude des performances de TCP dans un environnement réseau à très haut débit

Abdelbasset TRAD **Hossam AFIFI**

Projet PLANÈTE, Unité de recherche INRIA Sophia Antipolis

Résumé

Dans cet article nous présentons les résultats d'un ensemble d'expérimentations réalisées afin d'évaluer les performances des mécanismes de contrôle de flux du protocole TCP sur une connexion à très haut débit utilisant la fibre optique comme support de transmission.

Nous explorons, en premier lieu, l'effet de la bufférisation sur le débit de la connexion TCP. Nous identifions ensuite l'avantage de l'extension de la version TCP Reno qui utilise les acquittements sélectifs (TCP SACK) et qui permet de remédier à la perte de plusieurs paquets par fenêtre d'émission. Nous montrons aussi l'importance du mécanisme de mesure du RTT (Round Trip Time Measurement) qui utilise l'option TCP Timestamps afin d'éviter la dégradation de performance de TCP dans le cas d'un environnement réseau à très haut débit.

1 Introduction

La croissance soutenue que connaît le réseau Internet soulève de nombreux problèmes techniques parmi lesquels l'explosion du trafic transporté par les réseaux IP due à l'augmentation continue de la quantité de données et du nombre d'utilisateurs connectés à l'Internet.

L'introducion de la fibre optique comme support de transmission et de la technique de multiplexage en longueur d'onde WDM (Wavelength Division Multiplexing) qui exploite l'énorme bande passante disponible sur la fibre ont permis d'avoir un environnement réseau caractérisé par une vitesse de transmis sion très élevée qui répond à cette demande croissante en bande passante.

L'environnement réseau à haut débit a nécessité l'introduction de plusieurs améliorations [1,2] au protocole de transport TCP [4] responsable du contrôle de trafic et de congestion dans l'Internet afin d'adapter son fonctionnement à un tel environnement et de garantir sa fiabilité et sa performance.

Dans ce contexte le projet VTHD (Vraiment Très Haut Débit) [7] vise à anticiper les besoins de montée en débit des réseaux et à accélérer le développement des technologies de l'Internet Nouvelle Génération en déployant un réseau expérimental d'Internet à très haut débit dans le but de valider ou d'adapter les protocoles Internet de contrôle de flux et de congestion à l'environnement d'un réseau très haut débit ($>$1 Gigabit/s) et d'évaluer leur capacité à gérer des transferts de bout en bout à très haut débit .

Dans cet article on identifie dans une première partie les problèmes relatifs au fonctionnement de TCP dans le cas des réseaux haut débit et on décrit les améliorations introduites au protocole.

Une seconde partie est consacrée à la présentation des expérimentations qui ont été effectuées dans le but d'étudier les performances de TCP dans un environnement haut débit ainsi qu'à l'analyse des résultats obtenus.

Les tests ont été réalisés sur deux types de connexions utilisant la fibre optique comme support de trans-

mission:

1. Une connexion entre deux machines en point à point.
2. Une connexion de bout en bout sur le réseau expérimental VTHD.

2 Mécanismes de TCP

L'idée derrière les mécanismes de contrôle de flux et de congestion supportés par TCP est que chaque source de données contrôle sa vitesse de transmission indépendamment des autres de façon à réaliser les objectifs suivants:

- Utiliser au maximum la bande passante disponible sur les liens du réseau.
- Réduire le nombre de paquets en attente de transmission dans les routeurs.
- Distribuer la bande passante disponible de façon équitable entre les différents utilisateurs.

2.1 Algorithme Slow Start et Congestion Avoidance

L'algorithme Slow Start et Congestion Avoidance [3] introduit dans la version TCP Tahoe est du type basé sur fenêtre ("window based") et utilise une notification implicite de l'état de congestion dans le réseau. Cet algorithme permet à la source de contrôler la quantité de données soumise au réseau. Les variables introduites par cet algorithme sont les suivantes:

- ***cwnd*** (congestion window): Taille de la fenêtre de congestion, elle représente la quantité maximale de données que la source peut transmettre sur le réseau avant de recevoir un acquittement.
- ***ssthresh*** (slow start threshold): Seuil utilisé pour déterminer la phase de l'algorithme qui doit être exécutée (Slow Start ou Congestion Avoidance), généralement *ssthresh* est initialisé à la valeur de *rwnd* [1]

1. Receiver's advertised window: taille de la fenêtre d'émission indiquée par le récepteur.

Le ***min (cwnd, rwnd)*** est utilisé pour déterminer la quantité de données qui doit être transmise sur le réseau.

Slow Start

Après l'établissement d'une connexion ou après l'expiration d'un temporisateur de retransmission, l'émetteur fixe la taille de la fenêtre de congestion à 1 segment (MSS). A chaque réception d'un acquittement la taille de la fenêtre de congestion est incrémentée de 1 MSS:

$$cwnd = cwnd + MSS$$

Ceci revient à doubler la valeur de *cwnd* à chaque fois qu'une fenêtre de congestion entière est acquittée. Cet algorithme continue jusqu'à ce que la fenêtre de congestion atteigne le seuil "ssthresh".

Congestion Avoidance

Après la phase de Slow Start, qui prend fin au franchissement du seuil *ssthresh,* la fenêtre évolue de façon linéaire:

$$cwnd = cwnd + MSS * \frac{MSS}{cwnd}$$

Ceci revient à incrémenter la fenêtre par un MSS à chaque fois qu'une fenêtre de congestion entière est acquittée. Dans cette phase la croissance est linéaire plutôt qu'exponentielle pour éviter de revenir trop rapidement à une phase de congestion. Cette phase continue jusqu'à détection de perte de paquet à ce point TCP met à jour la valeur du seuil *ssthresh* à la moitiée de la fenêtre de congestion [6].

Le fait de passer à chaque perte d'un paquet la taille de la fenêtre à un segment, gaspille de la bande passante. De plus, une détection plus rapide des pertes de paquets permet d'avoir un débit plus élevé. Ces deux constatations ont permis d'implémenter de nouveaux mécanismes TCP: Retransmission Rapide et Récupération Rapide.

2.2 Algoritmes de Retransmission Rapide et de Récupération Rapide

Retransmission rapide (Fast Retransmit)

Pour avoir une indication plus rapide de la perte de paquets, l'idée d'acquittements dupliqués a été introduite [8], elle peut être décrite de la manière suivante:

À la réception d'un segment dont le numéro de séquence n'est pas celui attendu, le récepteur acquitte le dernier segment reçu dans l'orde. À la réception de quatre aquittements dupliqués l'émetteur retransmet le paquet perdu sans attendre l'expiration du RTO.

Recouvrement rapide (Fast Recovery)

C'est une optimisation de l'algorithme de retransmission rapide, ce dernier, suite à une détection de perte de paquets initialise la fenêtre de congestion à 1 et rentre dans la phase Slow Start souvent susceptible de baisser les performances de TCP. Le recouvrement rapide est un algorithme qui s'exécute juste après la retransmission rapide et qui permet d'éviter la phase Slow Start en passant directement à la phase Congestion Avoidance.

L'algorithme de retransmission rapide est intégré dans la version Tahoe et Reno de TCP. Le second, l'algorithme de recouvrement rapide, est intégré dans la version Reno uniquement et constitue l'unique changement par rapport à version Tahoe.

3 TCP et les réseaux haut débit

Plusieurs études ont montré que TCP [8] peut fonctionner de manière fiable et assez performante sur différents types de liens Internet dont la capacité de transmission de données varie de quelques centaines de bits/s (connexion Modem à 300 bits/s) jusqu'à l'ordre de centaines de Mbits/s, cependant plusieurs problèmes liés à la performance de TCP apparaissent dans le cas des réseaux à grand produit délai-bande passante. Ce paramètre est le produit de la bande passante[2] (bits/s) et du délai entre l'émetteur et le récepteur (RTT), il présente la quantité de données émises et non encore acquittées permettant d'avoir une utilisation efficace de la bande passante disponible [2].

Si ce produit est supérieur à 10^5 bits le réseau est qualifié de réseau à grand produit délai-bande passante. Pour le cas d'un réseau de fibre optique ayant un délai de 10 ms et une bande passante de 300 Mb/s ce produit dépasse 10^6 bits.

L'introduction de la fibre optique a permis d'atteindre des vitesses de transmission très élevées sur les liens, et vu que l'implémentation de TCP ne prenait pas en considération de tels aspects des problèmes apparaissent dans le cas des réseaux haut débit.

Une description de ces problèmes ainsi qu'un ensemble de solutions qui ont été introduites au protocole seront présentés dans cette partie.

3.1 Limitation de la taille de la fenêtre TCP

La vitesse de transmission élevée pour les réseaux haut débit est obtenue grâce à la quantité de données énorme pouvant transiter sur le lien (grande capacité du lien), toutefois l'entête TCP utilise un champ de 16 bits pour indiquer la taille de la fenêtre de réception (rwnd) ce qui limite la taille maximale de la fenêtre à 65 Ko (2^{16}).

Une nouvelle option de l'entête TCP ("window scale option") a été définie [2] afin de permettre d'avoir des fenêtres d'émission de taille plus grande, cette option définie un facteur multiplicatif pouvant avoir une valeur maximale de 14.

La taille maximale de la fenêtre de TCP pouvant être atteinte est alors de 2^{30} c'est à dire 1Go.

3.2 Rétablissement après une perte de plusieurs paquets

La perte de paquets dans un réseau à grand produit délai-bande passante résulte en une dégradation remarquable du débit d'une connexion TCP vu que cette perte entraîne la réduction de la fenêtre

2. La vitesse de transmission la plus élevée pouvant être atteinte, déterminée par la vitesse de l'élement le plus lent sur un chemin du réseau.

d'émission à la moitié et l'exécution de la phase "Slow Start". L'introduction des algorithmes "Fast Retransmit" et "Fast Recovery" [3] a permis le rétablissement de TCP après la perte d'un seul paquet par fenêtre d'émission, cependant l'expansion de la taille de la fenêtre d'émission augmente la probabilité de perte de plusieurs paquets par fenêtre.

Pour remédier à ce problème un mécanisme d'acquittements sélectifs (SACK) a été proposé [5].

TCP SACK

Le principe des acquittements sélectifs (SACK) consiste à reporter dans les acquittements les blocs contigus de segments reçus et non encore acquittés, l'émetteur retransmet alors les segments n'ayant pas été reçus en se référant aux intervalles de numéros de séquences vides entre les blocs indiqués.

L'implémentation SACK de TCP préserve l'algorithme de base de TCP Reno, la différence est seulement au niveau de la réaction à la perte de plusieurs paquets par fenêtre de données.

3.3 Estimation du RTO

L'implémentation de TCP garantie un transfert fiable de données en retransmettant les segments n'ayant pas été acquittés pendant un intervalle de temps RTO (Retransmission Time Out). Une estimation dynamique et précise du temporisateur de retransmission est déterminante pour la performance de TCP puisqu'elle permet d'éviter des retransmissions inutiles causées par une sous-estimation du RTO ou un retard de la détection de perte de paquets causé par une sur-estimation du RTO.

La valeur du RTO est déterminée en fonction de la moyenne et de la variance du RTT (Round Trip Time) qui est l'intervalle de temps entre l'émission d'un segment et la réception d'un acquittement de ce segment.

Plusieurs implémentations de TCP effectuent les mesures du RTT en utilisant comme échantillon un seul paquet par fenêtre, ceci permet d'avoir une approximation adéquate du RTT pour des fenêtres de taille réduite. Pour une vitesse de transmission très élevée une telle mesure engendre une erreur d'estimation du RTT vu la fréquence d'échantillonnage faible par rapport à la fréquence des données.

Mesures du RTT

Afin d'avoir une estimation plus précise du RTO pour les réseaux haut débit, un mécanisme de mesure du RTT (Round Trip Time Measurement) qui utilise une nouvelle option de TCP: *estampille de temps* (TCP timestamps option) a été introduit [1, 2]. Ce mécanisme permet de mesurer les RTT en effectuant la différence entre la valeur de l'estampille de temps envoyée dans un segment et la valeur de l'estampille reçue dans l'acquittement de ce segment.

4 Expérimentations

4.1 Tests en point à point

Afin d'évaluer les performances de TCP sur une connexion à très haut débit et d'identifier les paramètres limitatifs du fonctionnement du protocole dans un tel environnement un ensemble de tests a été effectué entre deux machines liées en point à point avec une fibre optique.

Les mesures ont été réalisées en utilisant les couches de protocoles natives de la plupart des Unix: l'implémentation des sockets sur TCP/IP.

4.1.1 Architecture de la connexion testée

Les deux machines pour lesquelles les tests ont été réalisés sont connectées en point à point sur des cartes réseau Gigabit Ethernet d'une capacité de transmission de 1Gbit/s.

L'architecture correspondante est représentée par la figure suivante:

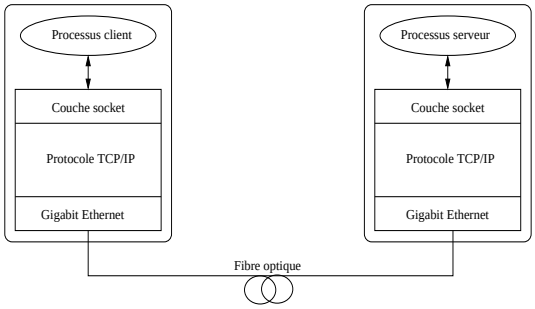

Figure 1 - Architecture de la connexion considérée

4.1.2 Effet de l'augmentation de la taille du buffer de la socket TCP

Cette expérience a pour but de mesurer le débit de transmission de données en TCP entre les deux machines considérées selon l'architecture représentée par la figure 1, et ceci en fonction de la taille des tampons (buffers) de la socket TCP.

La variation de taille de la fenêtre TCP entraînée par la variation de la taille des buffers de la socket est utilisée pour mettre en évidence les limites de performances de TCP.

On a essayé de maximiser le débit de transmission de données en augmentant la taille du buffer de la socket TCP, ceci en modifiant les paramètres *rmem_max* et *wmem_max* se trouvant respectivement dans les fichiers:

/proc/sys/net/core/rmem_max

et */proc/sys/net/core/wmem_max* sous Linux.

Ces paramètres représentent respectivement la taille mémoire réservée pour la réception et l'émission des données par la socket TCP.

Les résultats obtenus sont récapitulés dans le tableau suivant:

Taille des buffers	64 Ko	512 Ko	1 Mo	1 Go
Débit maximal	310 Mb/s	433 Mb/s	445 Mb/s	447 Mb/s

Tableau 1 - Débits mesurés en fonction de la taille des tampons de la socket TCP

D'après cette expérience on constate que le débit maximal pouvant être atteint en augmentant la taille du buffer jusqu'à une valeur de 1 Go est de 447 Mb/s.

4.1.3 Effet de l'estampille de temps TCP

Le but de cette expérience est de mettre en évidence l'impact de l'utilisation de l'estampille de temps de TCP introduite par l'option "TCP timestamps" [1, 2] sur le débit de la connexion considérée. Pour cela on a désactivé cette option en affectant la valeur 0 à la variable *tcp_timestamps* se trouvant dans le fichier */proc/sys/net/ipv4/tcp_timestamps* sous Linux (cette variable étant activée par défault). Le résultat est décrit dans le tableau suivant:

Option tcp_timestamps	activée	désactivée
Débit maximal	447 Mbits/s	477 Mbits/s

Tableau 2 - Débits mesurés en fonction de l'état de l'option tcp_timestamps de TCP

On constate d'après cette expérience qu'un gain de 30 Mbits/s apparaît sur le débit par rapport aux tests effectués avec l'option tcp_timestamps activée.

4.1.4 Effet des acquittements séléctifs

De la même manière que l'expérience précédente on a désactivé l'option TCP SACK en mettant la variable *tcp_sack* se trouvant dans le fichier:

/proc/sys/net/ipv4/tcp_sack sous Linux, à la valeur 0 (cette variable est activée par défault).

On a constaté que la désactivation de l'option TCP SACK n'a pas eu d'effet sur le débit de la connexion considérée, celui-ci est resté inchangé par rapport au débit engendré avec utilisation de SACK.

4.1.5 Analyse des résultats obtenus

Impact de la taille des buffers de la socket

La connexion qui a été testée dans cette partie est caractérisée par un délai de transmission très réduit ($250\mu s$) vu que les deux machines considérées sont directement liées sur les cartes Gigabit Ethernet par la fibre optique.

La valeur réduite du RTT implique une valeur du produit délai-bande passante peu élevée, c'est pour cette raison que le débit maximal a été atteint pour

une valeur du buffer de 512 Ko, et que la valeur du débit reste stable et inchangée suite à l'augmentation de la taille du buffer jusqu'à une valeur de 1Go (tableau 1).

Cependant le débit réduit obtenu pour une taille du buffer de 64 Ko peut être expliqué par la sous-utilisation de la bande passante de la fibre optique due à l'envoi d'une quantité de données inférieure au produit délai-bande passante.

Limitation de la valeur du débit

On constate que les débits engendrés pour la configuration considérée se limitent à 445 Mbits/s en TCP, ceci est principalement dû aux limitations du "hardware":

- Le bus PCI 32 bits fonctionnant à une vitesse d'horloge de 33 Mhz déjà limité physiquement à un débit de 132 Mo/s devient vite submergé pour les hauts débits.
- La taille de la mémoire RAM importe également, une taille de 512 Mo ne permet pas d'atteindre pleinement la bande passante de la fibre optique.

Effet de l'estampille de temps TCP

L'utilisation du mécanisme d'estampille de temps cause une réduction du débit de la connexion vu la charge induite pour l'émetteur et le récepteur (consommation de ressources système). L'emetteur maintient la valeur de l'estampille envoyée dans un segment et calcule la valeur estimée du RTT en fonction de l'estampille de temps envoyée par le récepteur dans l'acquittement de ce segment.

Cependant la désactivation de l'option TCP Timestamps peut engendrer de grandes dégradations de la performance de TCP qui ne peuvent pas être mises en évidence dans le cas de la connexion considérée vu le taux de perte très réduit qui la caractérise.

Effet des acquittements sélectifs

L'utilisation du mécanisme des acquittements sélectifs SACK n'a pas prouvé son intérêt dans le cas de la connexion considérée vu que le débit mesuré en désactivant l'utilisation de l'option TCP SACK ne diffère pas de celui obtenu en permettant l'utilisation de cette option.

Ceci peut être considéré comme étant un résultat logique vu que le mécanisme TCP SACK permet de remédier à la dégradation de l'utilisation de la bande passante suite à la perte de plusieurs paquets par fenêtre de transmission alors que dans le cas de l'environnement de la connexion testée le taux de perte de paquets est assez réduit.

4.2 Tests effectuées sur le réseau VTHD

Dans le but d'évaluer les performances de TCP sur une connexion de bout en bout à très haut débit, deux machines ont été connectées sur des cartes réseau Gigabit au réseau expérimental VTHD:

- Une machine locale se trouvant à l'INRIA Sophia Antipolis.
- Une machine distante placée à l'INRIA Paris.

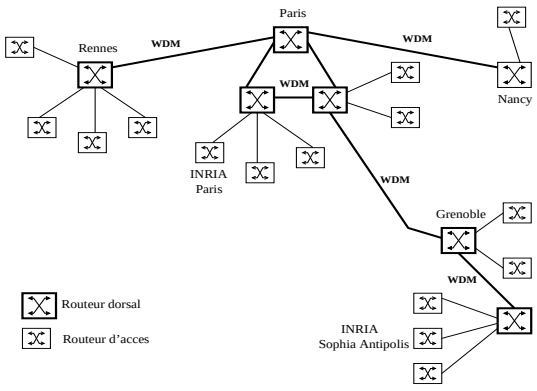

Figure 2 - Réseau expérimental VTHD testé

À la différence de la connexion en point à point testée dans la partie précédente, cette connexion est caractérisée par un délai de transmission plus grand et par la présence de quatre routeurs dorsaux, exploitant les capacités du multiplexage en longueur d'onde WDM, sur la liaison connectant les deux machines.

4.2.1 Description du réseau VTHD

Architecture physique
Le réseau dorsal VTHD comporte huit routeurs interconnectés par des canaux optiques à 2,5 Gbit/s. Sur ce réseau dorsal sont raccordés 18 routeurs d'accès agrégeant le trafic généré par les réseaux locaux d'entreprise des organisations exploitant la plateforme VTHD.

Système de transmission du réseau
Une architecture IP/WDM où les routeurs sont directement interconnectés par des systèmes de transmission optiques WDM est adoptée pour la partie dorsale du réseau VTHD. En associant sur une même fibre optique, plusieurs porteuses optiques ou longueurs d'onde, la capacité de transmission des fibres déployées est de l'ordre du térabit/s (10^{12}bit/s). Les longueurs d'onde interconnectant les routeurs du réseau dorsal VTHD portent chacune un débit de 2,5 Gbit/s.

Les liens d'accès interconnectant les routeurs dorsaux et les routeurs agrégeant le trafic des réseaux locaux utilisent le même principe consistant à transporter directement les flux de paquets IP sur les porteuses optiques. La technologie Ethernet dans sa version haut débit: Gigabit Ethernet, est utilisée sur les réseaux locaux. Les fibres d'accès n'exploitent pas cependant les capacités du multiplexage en longueur d'onde. Chaque site peut ainsi générer un débit de l'ordre du gigabit/s vers le réseau dorsal.

4.2.2 Mesures du débit de transmission de bout en bout

Dans cette expérience on mesure le débit d'une connexion TCP de bout en bout.

La connexion haut débit du réseau VTHD considérée présente les caractéristiques suivantes:

- une bande passante théorique de 1 Gbits/s disponible sur la fibre optique
- un délai de transmission moyen de 15 ms.

ce qui implique un produit délai-bande passante assez élevé (15 Mbits).

Pour mettre en évidence l'influence de la quantité de données transmise sur le débit de la connexion ayant un grand produit délai-bande passante, on a procédé à la variation de la taille des buffers d'émission et de réception de la socket TCP en modifiant à chaque fois les valeurs se trouvant respectivement dans les fichiers:

/proc/sys/net/core/wmem_max

et */proc/sys/net/core/rmem_max* sous le système Linux.

Les valeurs du débit obtenues sont récapitulées dans le tableau suivant:

Taille des buffers	64Ko	128Ko	512Ko	1Mo	512Mo	1Go
Débit (Mbits/s)	21,5	43,1	182	267	279	320

Tableau 3 -Débit en fonction de la taille des tampons de la socket TCP

D'après ce tableau on remarque bien que pour des buffers de taille inférieure à 1 Mo les débits engendrés sont relativement faibles par rapport à ceux obtenus pour 512 Mo ou 1 Go et qui dépassent les 270 Mbits/s. En passant de la taille du buffer prise par défault par le système Linux qui est de 64 Ko à une taille de 1 Go un gain très important (300 Mbits/s) a été obtenu pour le débit de la connexion.

4.2.3 Effet de l'utilisation des estampilles de temps

Le mécanisme d'estampilles de temps de TCP a été introduit dans le but d'assurer la fiabilité et la performance du fonctionnement de TCP dans le cas des réseaux haut débit caractérisés par une vitesse de transmission des données très élevée et ceci en permettant une estimation précise et dynamique du RTO à partir du calcul des RTT ce qui permet d'éviter le gaspillage de la bande passante causé par la retransmission inutile de paquets.

Afin de mettre en évidence l'effet de l'utilisation de cette option on a procédé à sa désactivation en exécutant la commande:
echo 0> /proc/sys/net/ipv4/tcp_timestamps sous Linux.

On a obtenu la courbe de débit suivante:

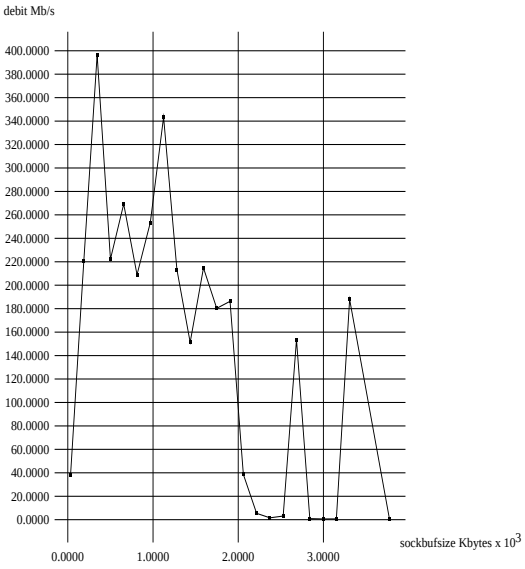

Figure 3 - Courbe du débit de TCP avec désactivation de l'option TCP Timestamps

Il est bien clair à partir de cette courbe que la performance de TCP se dégrade de manière remarquable lorsque l'option d'estampilles de temps de TCP n'est pas utilisée. Suite à l'augmentation de la taille des buffers de la socket TCP et en conséquent l'augmentation de la quantité de données transmises, le débit chutte de plus en plus vers des valeurs plus faibles pour atteindre des valeurs inférieures à 10 Mbits/s.

Cependant une valeur du débit qui dépasse les valeurs déjà obtenues lorsque l'option TCP Timestamps était activée a été obtenue (397 Mbits/s) mais ceci étant pour une taille du buffer de 300 Ko, c'est à dire pour un flux de données qui n'est pas très important.

4.2.4 Effet de l'utilisation du mécanisme TCP SACK

Dans le but d'identifier l'effet de l'utilisation des acquittements sélectifs sur le débit de la connexion TCP haut débit considérée une mesure du débit a été réalisée après désactivation de l'option TCP SACK par la commande:

echo 0 > /proc/sys/net/ipv4/tcp_sack.

La courbe du débit obtenue est la suivante:

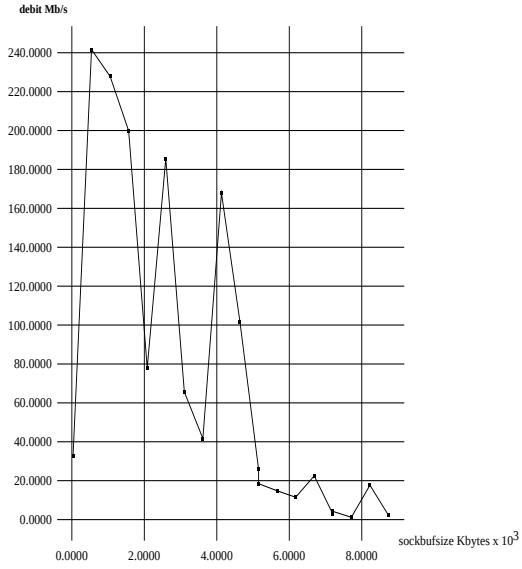

Figure 4 - Courbe du débit de TCP avec désactivation de l'option TCP SACK

D'après cette courbe la valeur maximale du débit est de 240 Mbits/s ce qui présente une diminution remarquable par rapport au débit maximal engendré losque l'option TCP SACK était activée.

On constate que le débit de la connexion diminue progressivement lorsqu'on augmente la taille du buffer de la socket pour atteindre des valeurs très faibles et voir inacceptables du débit ($<$ 10 Mbits/s).

4.2.5 Analyse des résultats de tests effectués sur le réseau VTHD

Effet de la bufferisation

La bande passante disponible sur le lien réseau qui est de 1 Gbits/s et le délai de transmission moyen qui est de 15 ms font que le produit délai-bande passante soit élevé (15 Mbits), ce qui nécessite une taille du buffer de la socket supérieure à 1,87 Mo afin de remplir le canal de transmission et de maintenir une utilisation efficace de la bande passante.

C'est pour cette raison que pour une taille du buffer de 64 Ko, qui est la taille prise par défaut par le système Linux, le débit est assez faible (21,5 Mbits/s).

Dans ce cas la source n'arrive pas à utiliser pleinement la bande passante disponible sur le lien vu la limitaion de la taille du buffer. Ce problème a pu être résolu en augmentant la taille du buffer de la socket TCP sous Linux, le débit atteint alors les 320 Mbits/s pour un buffer de 1 Go (tableau 3).

Effet des estampilles de temps

L'expérience réalisée en désactivant l'option d'estampille de temps de TCP prouve l'importance de l'utilisation de ce mécanisme dans le cas d'un transfert de données en haut débit, à l'opposé de ce qui a été constaté pour le test réalisé en point à point où l'utilisation de l'estampille de temps a montré un effet négatif sur la performance de TCP.

Ceci est dû au fait que le taux de perte de paquets croît en fonction de l'augmentation de la quantité de données envoyée par la source, les routeurs ayant des capacités de bufferisation limitées deviennent surchargés et par conséquent rejettent des paquets.

Le temporisateur de retransmission est un paramètre qui influe énormément sur les performances de TCP: si le RTO est sous-estimé TCP peut retransmettre inutilement les des segments déjà reçus et s'il est sur-estimé une perte sera détectée tardivement résultant en un temps d'inactivité de TCP assez important. Le mécanisme d'estampilles de temps s'avère primordial pour que TCP puisse se rendre compte rapidement de la perte de paquet et pour qu'il ait des informations précises et dynamiques reflétant l'état du réseau.

Avantage des acquittements sélectifs

L'expérience effectuée en désactivant l'option TCP SACK a permis de mettre en évidence le grand avantage de l'utilisation des acquittements sélectifs de TCP dans le cas d'une connexion haut débit. Lorsque les acquittements sélectifs ne sont pas utilisés le débit de la connexion se dégrade de plus en plus qu'on augmente la taille de la fenêtre d'émission. Cet effet peut être expliqué par l'augmentation de la probabilité de perte de plusieurs paquets par fenêtre d'émission losque la taille de celle ci est importante, la perte de paquets entraîne la réduction de la taille de la fenêtre d'émission à la moitié et l'exécution de la phase slow start. Les acquittements sélectifs permettent à TCP de se remettre de la perte de plusieurs paquets en évitant la grande réduction de la fenêtre au cas de l'exécution de la phase "slow start" et par conséquent permet de maintenir une utilisation stable et efficace de la bande passante du réseau.

5 Environnement des tests

Les tests ont été réalisés sur deux machines présentant la configuration matérielle suivante:

- Processeur: Pentium III, 933 Mhz.
- Bus: PCI 32 bits, 33 Mhz.
- Carte réseau: Carte 3Com 3C985 SX Gigabit Ethernet.
- RAM: 512 Mo.

Les tests ont été effectués sur un noyau Linux 2.2.16.

L'outil NTTCP (New Test TCP program) version 1.47 [9] a été utilisé pour effectuer les mesures de débit, celui-ci est lancé en mode réception sur une machine et en mode émission sur une autre transmet un nombre déterminé de buffers de taille définie en mesurant le débit de la connexion utilisée.

6 Conclusion

Les mécanismes de contrôle du flux de bout en bout du protocole TCP révèlent leur robustesse et leur adaptabilité à l'environnement d'un réseau à très haut débit vu l'implémentation de TCP qui ne considère pas d'hypothèses sur le type d'infrastructure ou d'équipements réseaux utilisés.

D'après les expérimentations effectuées sur le réseau haut débit VTHD on peut souligner l'importance du mécanisme d'estampilles de temps de TCP dans le cas d'un réseau à très haut débit qui nécessite une estimation précise du temporisateur de retransmission afin d'éviter le gaspillage de bande passante suite à des retransmissions inutiles causées par une mauvaise estimation du RTO.

Les performances de TCP sont nettement meilleures lorsque le mécanisme d'acquittements sélectifs est utilisé, vu la probabilité élevée de perte de plusieurs paquets par fenêtre dans le cas d'un réseau haut débit caractérisé par une fenêtre d'émission de taille importante. Les acquittements sélectifs permettent d'éviter la dégradation de performance de TCP causée par l'expiration du temporisateur de retransmission et l'exécution de la phase "slow start" suite à la perte de plusieurs paquets.

La taille du buffer de la socket TCP influe également sur l'utilisation de la bande passante. Pour une bande passante de 10 Mbits/s ou 100 Mbits/s un buffer de 64 ko suffit, alors qu'en Gbits/s on s'aperçoit qu'un buffer d'au moins 1 Mo est indispensable.

L'augmentation de la taille du buffer de la socket TCP permet une utilisation plus efficace de la bande passante disponible sur la fibre optique (débit maximal de 320 Mbits/s en TCP) cependant des limitations liées à l'architecture matérielle (bus, mémoire) surviennent et font que la bande passante ne soit pas pleinement utilisée.

Références

[1] RFC 1185, TCP Extention for High-Speed Paths. V. Jacobson, R. Braden, L. Zhang. Network Working Group, Octobre 1990.

[2] RFC 1323, TCP Extentions for High Performance. V. Jacobson, R. Braden, D. Borman. Network Working Group, Mai 1992.

[3] RFC 2001, TCP Slow Start, Congestion avoidance, Fast Retransmit, and Fast Recovery Algorithms. W. Stevens, Network Working Group, Janvier 1997.

[4] RFC 793, Transmission Control Protocol, spécifications du protocole. Defense Advanced Research Projects Agency, Septembre 1981.

[5] RFC 2018, TCP Selective Acknowledgment Options. M. Mathis, J. Mahdavi, S. Floyd, A. Romanov. Network Working Group, Octobre 1996.

[6] RFC 2581, TCP Congestion Control. M. Allman, V. Paxon, W. Stevens. Network Working Group, Avril 1999.

[7] Projet VTHD: Vraiment Très Haut Débit. France Télécom, INRIA, groupe des écoles de télécommunications. http://www.vthd.org

[8] Van Jacobson, Congestion Avoidance and Control. Actes de ACM SIGCOMM'88, Standford, 1988.

[9] Elamar Bartel, NTTCP. Université de Munchen, Octobre 1998. http://home.leo.org/~bartel/nttcp/nttcp.