

帯域予約されたTCPトラフィックの特性評価

久田 裕介 小野 諭

hisada@slab.ntt.co.jp ono@slab.ntt.co.jp

NTTソフトウェア研究所

本稿では、実測定に基づいて、TCPトラフィックを帯域予約して送信した場合のスループットを評価した。その結果、端末-端末間のラウンドトリップ遅延が帯域予約によるフレーム送信間隔よりも短い場合には、TCPトラフィックのスループットが低下することと、予約帯域を回線帯域と等しくした場合にはスループットが急に高くなることを明らかにした。評価結果をもとにして、これらのスループット特性は、レート制御が入出力待ち行列を分離、連結することによって変化することを指摘した。

Performance Evaluation of Bandwidth Reserved TCP Traffic

Yusuke HISADA Satoshi ONO

hisada@slab.ntt.co.jp ono@slab.ntt.co.jp

NTT Software Laboratories

We evaluate TCP performance over a bandwidth reserved link. Results based on measurements show considerable variations in throughput. First, We show that average throughput become lower as end-to-end round trip delay become shorter than frame interval, because a rate control mechanism we evaluates splits input queue and output queue. Second, we show that average throughput depends on link capacity even if same bandwidth is reserved, because the rate control mechanism concatenates input queue and output queue.

1 はじめに

ベストエフォートの通信における帯域、遅延は他のトラフィックの影響によって変動するため、通信品質を保証するために資源予約が行われている。一方、誤りのない通信をするためには、TCPが広く用いられている。

TCPの性能評価は多様な観点から行われている[3][4]。特に、リーキーパケット、トークンパケット等のレート制御方式と、優先度制御、重み付け公平待ち行列法(Weighted Fair Queueing: 以下、WFQと略す)等のスケジューラを用いた性能保証の研究では、トラフィックを最初にネットワークに入力する際にリーキーパケット方式とトークンパケット方式を用いてレート制御し、WFQでスケジューリングすれば、端末-端末間の遅延上限を保証することが

証明されている[1]。しかしながら、この遅延上限は一般に長いため、遅延分散を十分に抑えることはできない。

一方、WFQの代わりに優先度制御を用いて、高優先度の待ち行列のTCPを割り当てれば、その遅延分散を抑えることができる。しかしながら、優先度制御では低優先度フローの遅延上限を保証しないため、高優先度フローはバースト的に送信される。このため、レート制御と組み合わせると、フレームが大量に廃棄される。フレーム廃棄はTCPの性能劣化を引き起こすため、多段で優先度制御を実行するのは適切ではない。

従って、遅延変動を抑えるためには、レート制御、WFQ、優先度制御のすべてを用いる必要がある。ただし、これらを独立して実行すると、上

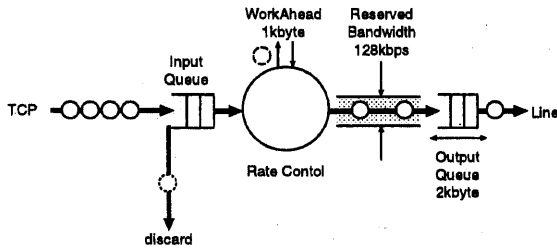


図1 本実験におけるレート制御方式

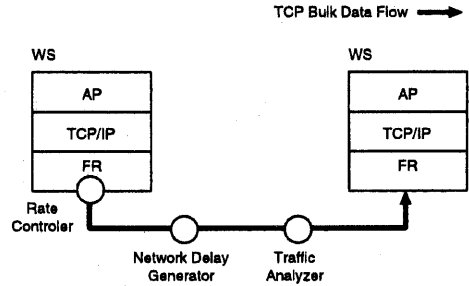


図2 測定

記機能間で待ち行列を用意しなければならないため、遅延変動を大きくする恐れがある。このため、各機能を統合的に実行することも必要である。これらの問題意識に基づいて、筆者らは、先に、要求駆動型の帯域予約方式[2]を提案した。

この方式は、既存の各種レート制御技術、各種スケジューリング技術を共通パラメータで制御するためのフレームワークである。従って、これを用いて評価することにより、従来のレート技術を用いてTCPの帯域予約を行う場合の問題を抽出することができると考えられる。そこで、本稿では、要求型多重化方式のレート制御方式のTCPのスループットを、実測定を通して評価する。

本稿では、第2節で実験について説明し、第3節では測定結果を評価する。本稿におけるレート制御方式では、ラウンドトリップ遅延がフレーム送信間隔よりも短いとスループットが低下する。また、予約帯域をリンクレートに近づけるとスループットが高くなる。第5節ではスループットが変化した場合の原因を考察する。本実験のレート制御方式では、ラウンドトリップ遅延がフレーム送信間隔よりも短いとレート制御への入力待ち行列でオーバーフローが発生しやすい。また、予約帯域とリンクレートを等しくすると、レート制御への入出力待ち行列のフレーム送受信が同期するため、オーバーフローが発生しにくくなる。

2 実験

本実験におけるレート制御を図1に示す。本実験では、単位時間あたりの送信データ量を一定に保つようにフレーム送信間隔を制御する。また、1kbyteまでのデータをワークアヘッド分として送信するように制御する。なお、入力待ち行列に入ることのできないデータは廃棄される。

本実験の装置構成を図2に示す。前述のレート制御方式は、端末(Sun SPARC Station 10, SunOS4.1.3)のフレームリレー通信ボード上にハー

ドウェアとドライバを用いて実装する。端末間の中継回線はラインレート1.536Mbpsのインタフェース専用線で、128, 192, 384, 768kbpsの中継回線としても利用する。また、端末-端末間の網内遅延を変えるための網内遅延発生装置と、フレーム通過時刻測定装置を経路に挿入している。

3 評価

本節では、中継回線のリンクレート、網内遅延、入力待ち行列容量を変えた場合の平均スループットを評価した。ここでいう平均スループットとは、最初のデータフレームが測定点を通過してから、最後のデータフレームがその測定点を通過するまでに送信されたデータ量の時間平均値である。

3.1 網内遅延を変えた場合

リンクレート、入力待ち行列容量を1.536Mbps, 2kbyteに固定し、網内遅延を変えた場合のスループットを図3, 図4に示す。図3では、網内遅延30ミリ秒における平均スループットが最も高く、網内遅延が長くなるほど平均スループットは低くなる。これは、網内遅延が長くなることにより、1回のフレーム廃棄に伴うTCPの再送タイムアウト時間も伸びるためである。

一方、図4の場合、上とは逆に、網内遅延が30ミリ秒における平均スループットが最も高く、網内遅延が短くなるほど平均スループットは低くなる。次に、リンクレート、入力待ち行列容量を128kbps, 2kbyteに固定し、網内遅延を4ミリ秒から30ミリ秒の範囲で変えた場合のスループットを図5に示す。ここでも、網内遅延が30ミリ秒よりも短くなるほど、平均スループットは低下する。これは、ラウンドトリップ遅延がフレーム送信間隔よりも短くなったために発生したと考えられる(第4.2節)。

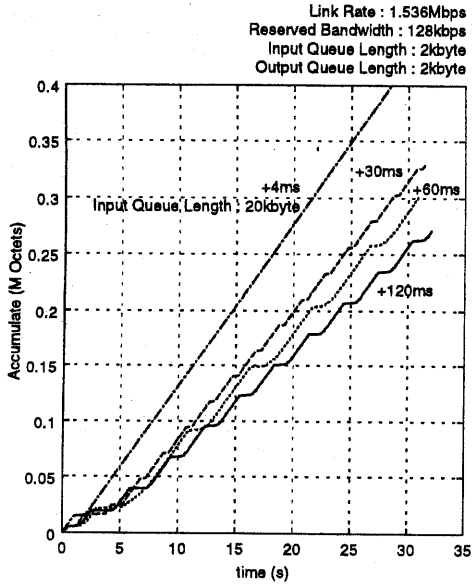


図3 網内遅延を変えた場合のスループット(リンクレート1.536Mbps, 予約帯域128kbps, 入力待ち行列容量2kbyte)

3.2 リンクレートを変えた場合

網内遅延, 入力待ち行列容量を4ミリ秒, 2kbyteに固定し, リンクレートを変えた場合のスループットを図6に示す. リンクレートが384kbpsより低い場合には, リンクレートに対する予約帯域の割合が高くなるほど, 平均スループットは高くなる. 一方, リンクレートが384kbps以上の場合には, 平均スループットに大きな変化は見られない. また, 本実験のレート制御方式では, 予約帯域がリンクレートに等しくなると平均スループットは急に高くなる. これは, 入力待ち行列ラウンドトリップ遅延がフレーム送信間隔と等しくなったために発生したと考えられる(第4.2節).

3.3 入力待ち行列容量を変えた場合

リンクレート, 網内遅延を1.536Mbps, 4ミリ秒に固定し, 入力待ち行列容量を変えた場合のスループットを図7に示す. 本実験で用いたレート制御方式では, 入力待ち行列容量を大きくすると, 平均スループットは高くなる. これは, 入力待ち行列容量を大きくするとオーバーフローの可能性が小さくなるためと考えられる.

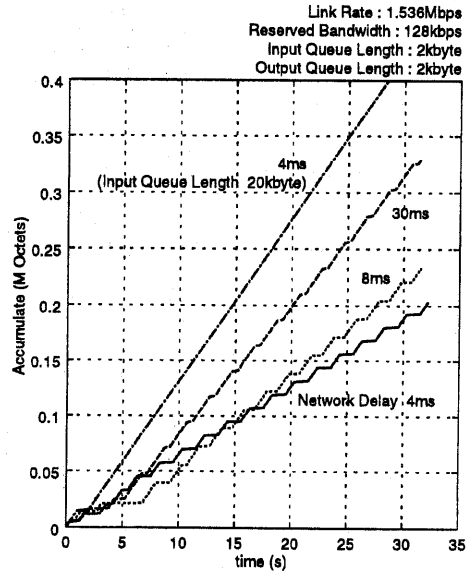


図4 網内遅延を変えた場合のスループット(リンクレート1.536Mbps, 予約帯域128kbps, 入力待ち行列容量2kbyte)

3.4 スループットの最悪値

前節までの評価結果に基づいて, 各パラメータを最も厳しい条件に設定した場合, つまり, リンクレート, 網内遅延, 入力待ち行列容量を1.536Mbps, 4ミリ秒, 2kbyteにした場合のスループットを図6に示す. リンクレートに対する予約帯域の割合が低く, 網内遅延が短く, 入力待ち行列容量が小さい場合には, フレーム送信開始からしばらくの間, スループットが極めて低くなる(図8). これは, 再送中に重ねてフレーム廃棄が発生し, TCPの再送タイムアウト時間が最大8秒まで長くなったためである(第4.3節).

4 考察

本節では, まず, 第4.1節において, 網内遅延が短いとスループットが低下する現象(第3.1節)の原因を考察する. 次に, 第4.2節において, 予約帯域をリンクレートに近づけるとスループットが高くなる現象(第3.2節)の原因を考察する. 最後に, リンクレートを高く, 網内遅延を短く, 入力待ち行列容量を小さくすると, スループットが極めて低くなる現象(第4.3節)の原因を述べる.

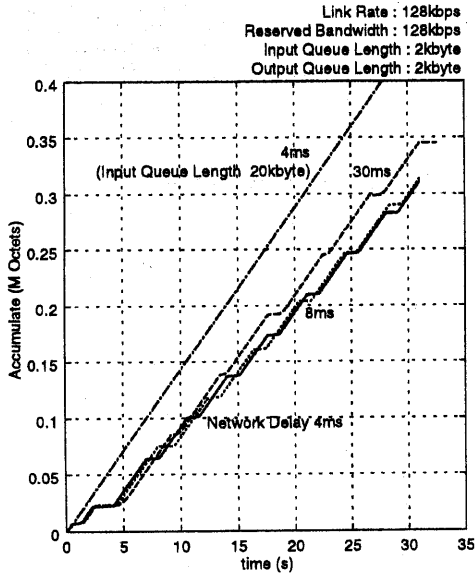


図5 網内遅延を変えた場合のスループット(リンクレート128kbps, 予約帯域128kbps, 入力待ち行列容量2kbyte)

4.1 ラウンドトリップ遅延とフレーム送信間隔

ラウンドトリップ遅延がフレーム間の送信間隔よりも短い場合には、TCPは、入力待ち行列からフレームが送信される前に、入力待ち行列から前回送信されたフレームへの確認応答を受信する。逆に、ラウンドトリップ遅延がフレーム間の送信間隔よりも長い場合には、TCPは、入力待ち行列からフレームが送信された後に、入力待ち行列から前回送信されたフレームへの確認応答を受信する。その結果、前者の場合における見かけの入力待ち行列容量は、後者の場合よりも小さくなる。

本実験のレート制御方式では、バースト的な送信を許可した期間以外には、フレーム送信間隔は32ミリ秒である。一方、網内遅延が4ミリ秒、8ミリ秒の場合には、ラウンドトリップ遅延はフレーム送信間隔よりも短いため、見かけの入力待ち行列容量も小さい。入力待ち行列容量が小さいほど平均スループットは低下するため(図7)、網内遅延4ミリ秒、8ミリ秒の場合には、平均スループットが低下すると考えられる(第3.1節および図4, 図5)。

4.2 予約帯域とリンクレート

予約帯域とリンクレートが等しい場合には、入力待ち行列から出力待ち行列へのフレーム送信時間と

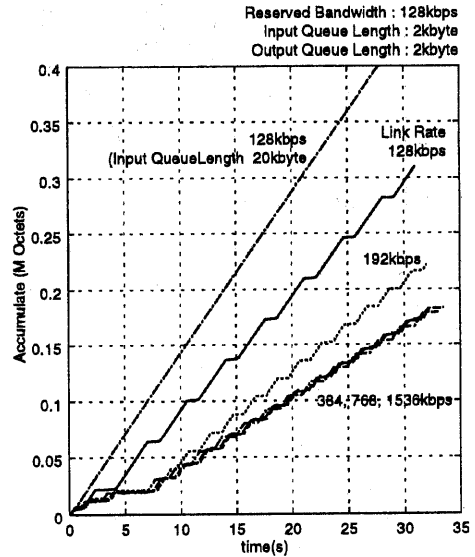


図6 リンクレートを変えた場合のスループット(予約帯域128kbps, 網内遅延4ms, 入力待ち行列容量2kbyte)

出力待ち行列から中継回線へのフレーム送信時間が一致する。このため、入力待ち行列の先頭フレームは、出力待ち行列が空いたら、直ちに出力待ち行列へ送信される。つまり、本実験のレート制御方式によって、入力待ち行列と出力待ち行列が1本の待ち行列として扱われる。その結果、入力待ち行列と出力待ち行列がともに満杯になるまで、フレームは廃棄されない。

一方、予約帯域がリンクレートよりも低い場合には、入力待ち行列から出力待ち行列へのフレーム送信時間が出力待ち行列から中継回線へのフレーム送信時間よりも長い。このため、入力待ち行列の先頭フレームは、出力待ち行列が空いても、直ちに出力待ち行列へ送信されない。つまり、本実験のレート制御方式によって、入力待ち行列と出力待ち行列は独立した2本の待ち行列として扱われる。その結果、出力待ち行列が空いていても、入力待ち行列が満杯になると、フレームは廃棄される。

以上より、予約帯域がリンクレートと等しい場合には、オーバーフローが発生しにくいと考えられる(第3.2節および図6)。

このような、入出力待ち行列間のフレーム送受信が同期するためにスループットが高くなる現象を、本稿では、待ち行列の連結(queue cat)と呼ぶ。逆

に、入出力待ち行列間のフレーム送受信が非同期なためにスループットが低くなる現象を、本稿では、待ち行列の分離(*queue split*)と呼ぶ。

4.3 再転送シーケンスのネスト

本実験で用いたTCPスタックは、最初のフレームを送信する際にスロースタートを実行しないため、TCPは、3過程ハンドシェイクの直後に合意したウィンドウサイズに等しい量のデータを送信する。一方、図8の測定条件では、予約帯域のリンクレートに占める割合が低い。従って、図8では、最初のフレーム送信時に14kbyte分のデータがまとめて廃棄され、その結果、TCPは非常に長い再送シーケンスに入っている。しかも、再送シーケンス中にさらにフレーム廃棄が発生したために再送タイムアウト時間が倍数増加し、その結果、平均スループットが極端に低下した。

5 まとめ

本稿では、TCPを帯域予約した場合のスループットを実測定に基づいて評価し、ラウンドトリップ遅延が短いとスループットが低下すること、予約帯域をリンクレートに近づけるとスループットが高くなることを明らかにした。これらの評価結果をもとに、スループットはレート制御が入出力待ち行列を分離、連結することによって変化することを指摘した。

謝辞 本研究の機会を与えていただいた、NTTソフトウェア研究所広域コンピューティング研究部の市川晴久部長、日頃ご指導いただき高橋直久グループリーダー、ならびに測定において多大なご支援をいただいた福田晴元さんに感謝します。

参考文献

- [1] A. K. J. Parekh, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks", LIDS-TH-2089, MIT Laboratory for Information and Decision Systems, Cambridge, MA, 1992.
- [2] 神林, 山下, 小野, "QoSを考慮した適応型多重化方式", 情報処理学会, マルチメディア通信と分散処理ワークショップ論文集, 1993.
- [3] Y. Ahek, Y. Mansour, and Z. Ostfeld, "Phantom: A Simple and Effective Flow Control Scheme", Computer Communication Review, Vol. 26, No. 4, pp. 169-182, October 1996.
- [4] 長谷川, 阿野, 加藤, "シミュレーションを用いた広域ATM網におけるTCPの性能評価", 情報処理学会第54回全国大会論文集, 1997.

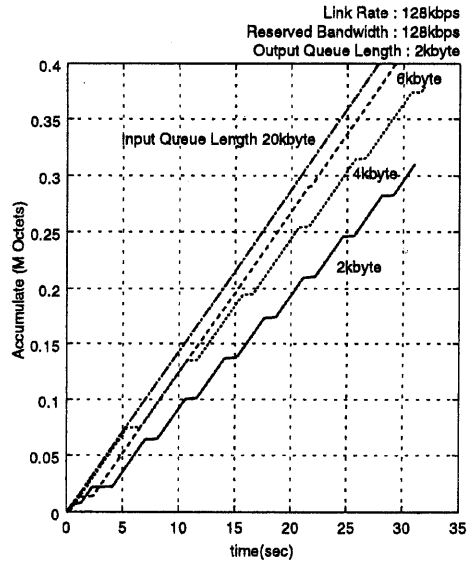


図7 入力待ち行列容量を変えた場合のスループット(リンクレート1.536Mbps, 予約帯域128kbps, 網内遅延4ms)

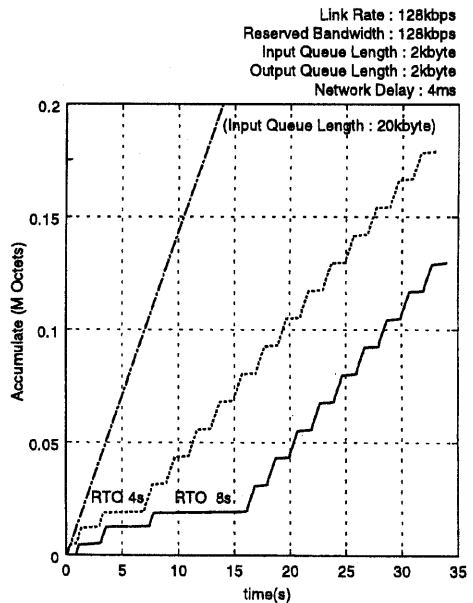


図8 スループットの最悪値(リンクレート1.536Mbps, 予約帯域128kbps, 網内遅延4ms)