TCPフロー制御における帯域共有のモデル

石橋　圭介，川原　亮一

本稿では，TCPのフロー制御における帯域配分がどのようにモデル化されるかを述べる．とくに単一ボトルネックリンクネットワークにおいてはプロセッサシェアリング待ち行列でモデル化できることを示し，その結果得られる性能指標について述べる．また，このモデルを実際のネットワーク環境に応じて拡張した結果についてもいくつか紹介する．

キーワード：TCP，プロセッサシェアリングモデル，公平性

1. はじめに

インターネットではデータ転送以外の機能を端末側に集約させるという思想（End-to-end principle）の下，ネットワークの輻輳制御は，主に端末上のTransmission Control Protocol（TCP）のフロー制御として実現されている．したがって，World Wide Webやファイル転送など，TCPを用いてデータを転送するアプリケーションの性能はTCPのフロー制御に依存して決まり，これらアプリケーションの性能評価のためにはTCPの性能評価が必要となる．TCP性能解析には，大きく以下の二つのアプローチがある．

1. パケット損失や遅延の発生により送信速度を調整するというTCPフロー制御のメカニズムを詳細にモデル化することで，パケット損失率やパケット遅延からスループットなどの性能を求める方法
2. TCPフロー制御の詳細なメカニズムには立ち入らず，TCPフロー制御が各コネクションに対して帯域を公平に配分するよう設計されている点に注目し，資源の配分モデルとして待ち行列モデルの一つであるプロセッサシェアリング待ち行列によりモデル化し，コネクションの発生率や転送データサイズから平均ファイル転送時間などを求める方法

前者は，より収集が簡易なパケットレベル性能から上位レイヤであるTCPの性能を推定するものであり，ユーザによる品質管理などに有用と思われる．一方後者はネットワークの条件をパラメータとしてTCP性能を得る事ができるため，ネットワーク設計や管理のために有用であると思われる．

本稿では後者について概説する．前者については文献[1, 2]などを参照されたい．後者は複数のコネクション間でボトルネックリンク帯域が公平かつ効率的に配分されるとしてモデル化するものであるが，どのように公平に配分されるのかをいったんパケットレベルに立ち戻って簡単に節2で述べる．節3で，その効率性および公平性によってTCPフロー制御における帯域配分がプロセッサシェアリング待ち行列でモデル化できることを述べる．節4で，このモデルをさらに実際のTCPの挙動に対応させて精密化する試みについていくつか紹介する．

2. TCPフロー制御 ―帯域の公平な配分―

TCPフロー制御の目的はリンク帯域の効率的な利用，およびリンク中に存在するコネクション間の公平な帯域配分である．しかしながら，節1に述べたように，TCPフロー制御は端末上に具備されており，経路上のボトルネックリンクの帯域，そのリンク中の他のコネクションの数などのネットワーク情報は利用できない．したがって，TCPフロー制御はこれらの情報なしに，適切な送信速度を推定し，効率的かつ公平な帯域配分を達成しなければならない．以下，ボトルネックリンク以外にスループット低下要因がないものと仮定して説明する．

TCPでは，送信端末がパケットを送信し，受信端末はそれに対して受信確認（Ack）を返信することによって信頼性のあるデータ転送を実現している．この

いしばし　けいすけ
NTT 情報流通プラットフォーム研究所
〒180-8585 武蔵野市緑町3-9-11
かわはら　りょういち
NTT サービスインテグレーション基盤研究所
〒180-8585 武蔵野市緑町3-9-11

とき送信端末は，送信パケットのAckを待たずに複数個のパケットを送信することができる．このパケット群の転送データ量はウィンドウサイズと呼ばれるが，これは送信端末がパケット-Ack往復時間の間に送信できるデータ量であり，したがってTCPのスループット（単位時間当たりの送信データ量，送信速度）はウィンドウサイズ W[bit] を往復遅延時間 D[s] で割った値 W/D とほぼ一致する．

TCPフロー制御はこのウィンドウサイズを調節することによって，データ送信速度を制御している．コネクション開始時にはウィンドウサイズを1パケット分から開始し，その後，Ackを受信するごとに徐々に大きくしていく．ウィンドウサイズは，コネクション毎に上限 W_M[bit] が決められ，スループットの上限はこれを往復遅延時間（正確には往復伝搬遅延）で割った値 W_M/D となるが，これがボトルネックリンク帯域 C[bps] より大きければ，最終的にこのコネクションはリンク帯域を使いきれることになる．

また，パケット損失が発生した場合，コネクションのデータ送信速度（の和）がリンク帯域を超えてしまい輻輳が発生したと判定して，いったんウィンドウサイズを縮小し，その後またウィンドウサイズを大きくしていく．したがって，コネクション間でパケット損失率が同じであり，かつ往復遅延時間も同一であるとすると，ウィンドウサイズの平均的挙動が一致するためにコネクション間のスループットも一致し，公平な帯域配分が得られることになる．

以上の仮定（コネクションの往復遅延時間が同一，単一ボトルネック等）のもとでは，ボトルネックリンクで複数のTCPコネクションが競合した場合，TCPフロー制御によってコネクション間でリンク帯域が効率的かつ公平に配分されることになる．より詳細なTCPフローの公平性については，文献[3]等を参照されたい．

3. プロセッサシェアリング待ち行列によるモデル化

プロセッサシェアリング（Processor sharing, PS）は，待ち行列におけるサービス規律の一種であり，待ち行列の先頭にいる客のみがサービスを受けるFirst Come First Served（FCFS）待ち行列と違い，システム内の客すべてにサーバ能力が公平に割り当てられるサービス規律である．前節の説明により，TCPコネクションはボトルネックリンク帯域を公平に分け合うので，TCPのフロー制御による帯域共有をPS待ち行列でモデル化できることになる[4]．すなわちコネクションを客として，コネクションの発生を客の到着とみなし，時刻tの同時接続コネクション数 $n(t)$ を系内客数として，$n(t)$ 人の客すべてが等しく $C/n(t)$ の処理能力でサービスされるとするモデル化である（したがって，キューイングなどのパケットレベルの挙動は一切無視されている）．また，要求サービス時間はコネクションが転送すべきサイズをボトルネックリンク帯域で割った値であり，系内滞在時間はコネクションの転送時間である．サーバ使用率 ρ はボトルネックリンク使用率であり，コネクション発生率 λ[1/s] と平均転送データサイズ s[bit] をもちいて $\rho=\lambda s/C$ とあらわされる．

PS待ち行列は，客がポアソン過程に従って到着するとき，M/M/1 FCFSと系内客数分布が一致するため，平均系内客数は $\rho<1$ のとき要求サービス時間分布（または転送データサイズ分布）によらず，

$$\mathrm{E}[n(t)]=\frac{\rho}{1-\rho} \tag{1}$$

であり，ファイル転送時間 T[s] の平均はリトルの式により

$$\mathrm{E}[T]=\frac{s}{C(1-\rho)} \tag{2}$$

という簡易な式で与えられる．

4. 精度を上げるためのモデルの拡張

節3でのべた結果は，TCPフロー制御における帯域共有を表現したモデル化に基づくものであった．このモデルを，TCPの挙動や，ネットワークの環境を反映させてさらに拡張することによって，性能評価の精度をあげる試みがなされている．本節ではそれらの内のいくつかを紹介する．

4.1 均一なアクセス帯域制限のあるモデル

節3ではボトルネックリンク内に単一コネクションのみ存在する場合は，そのコネクションがリンク帯域を使いきれると仮定した．しかし，ユーザのコネクションは対象とするボトルネックリンク以外にもより低速なアクセスリンクを経由している場合が考えられる（図1）[1]．このとき，たとえボトルネックリンク中の同時接続TCPコネクション数が1本であったとして

[1] この場合，着目しているリンクの非混雑時にボトルネックとなるのはアクセスリンクであり，当該リンクはボトルネックリンクでなく，集約リンクと呼ぶべきであるが，本稿ではボトルネックリンクで統一する．

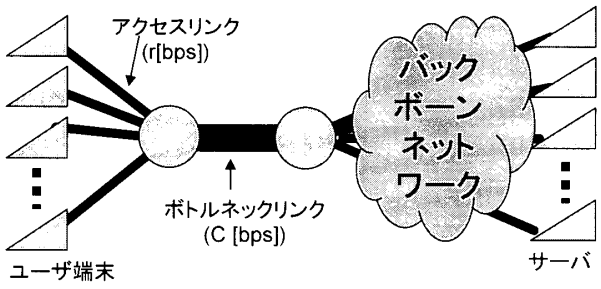

図1　アクセスリンク帯域制限ありの場合のネットワーク

も，そのコネクションはボトルネックリンク帯域 C [bps] を使い切ることはできない．

アクセスリンク帯域がすべて同一で $r(<C)$[bps] とすると，同時接続コネクション数が $n(t)$ 本のときの各コネクションの利用可能帯域は $\min(C/n(t), r)$ となる．すなわち，コネクション数 $n(t)$ が C/r 未満の場合は，アクセスリンク帯域ネックであり，コネクションの利用可能帯域は r[bps] となり，C/r 以上となると，ボトルネックリンク帯域ネックのため，利用可能帯域が $C/n(t)$[bps] に制限される．この場合は，処理能力 C[bps] の単一サーバPS待ち行列でなく，処理能力 r[bps] の C/r 個の複数サーバPS待ち行列でモデル化することができる[2] [5, 6]．このモデル化において，各コネクションがポアソン到着する場合，平均ファイル転送時間は転送ファイルサイズ分布に依らず次式で与えられる．

$$\mathrm{E}[T]=s\left(\frac{1}{r}+\frac{\mathbf{C}(\rho, C/r)}{C(1-\rho)}\right) \tag{3}$$

ここで $\mathbf{C}(\rho, C/r)$ はアーランC式に従い，リンク帯域が使い切られている（同時接続コネクション数が C/r 本以上である）確率を表す．したがって，ボトルネックリンク使用率 ρ が低い場合の平均転送時間は，ほぼアクセスリンク帯域で決まり，s/r で与えられるが，ボトルネックリンク使用率が高くなるにつれてボトルネックリンク帯域を使い切る確率が増大し，値は式(2)に近づくことになる．図2に，ボトルネックリンク使用率の増大に対する，複数サーバの場合（式(3)）と単一サーバの場合（式(2)）のTCP平均転送時間の増大度を比較する．前出の事象が確認できる．

4.2　不均一なアクセス帯域制限のあるモデル

節4.1では，アクセスリンク帯域がすべてのフローで同一と仮定したが，実際には，ボトルネックリンク中のコネクションは様々な環境のアクセスリンクから流入している．このような場合，アクセスリンクの帯域によって，ボトルネックリンク混雑時の平均転送時間の増え方は異なってくると考えられる．

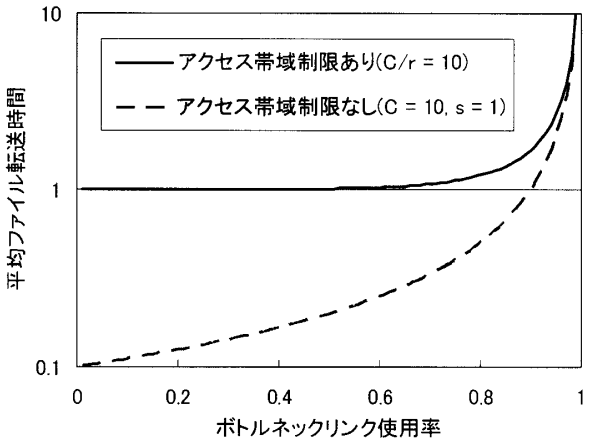

図2　アクセスリンク帯域制限ありとなしの場合の比較

このような状況下においては，各コネクションの利用可能帯域はmax-min公平性と呼ばれる帯域配分に従うと考えられる．max-min公平性とは，利用可能帯域が最小となるコネクションの利用可能帯域を最大化する公平性である[7]．例えば，9 Mbpsのボトルネックリンク上において，アクセスリンク帯域が1 Mbpsのコネクションが5本，アクセスリンク帯域が5 Mbpsのコネクションが2本，通信中であるとする．このとき，max-min公平性に従えば，まず，アクセスリンク帯域が1 Mbpsの各コネクションにそのアクセスリンク帯域分の帯域（=1 Mbps）が配分され，残りの（9 Mbps−1 Mbps×5本）=4 Mbpsが，アクセスリンク帯域が5 Mbpsのコネクション各々に対し，2 Mbpsずつ配分されることになる．

このような場合，ある瞬間のコネクションの利用可能帯域はボトルネックリンク中のコネクション数のみならず，アクセスリンク帯域毎のコネクション数に依存し，これらコネクション数の定常状態確率を閉形式で表現すること，もしくは平均転送時間を求めることは非常に困難である．文献[8]では，このような環境におけるコネクションの平均転送時間近似式として，式(3)における r を着目するコネクションのアクセスリンク帯域 r'[bps] で置き換えたものを提案している（式(3)における s も，着目するコネクションの平均ファイルサイズ s' で置き換える）．こうすることにより，着目するコネクションの特性 (r', s') とボトルネックリンク特性 (C, ρ) のみによって平均ファイル転送時間を計算できる．

この近似式では，着目するコネクション以外の他のコネクションも全てアクセスリンク帯域 r' を持つと

[2] 簡単のため，本稿では C は r の整数倍とする．

みなすことになる．もしコネクションへの帯域配分がmax-min公平性に従っているとすると，ボトルネックリンクが輻輳している場合には，他のコネクションのアクセス帯域の値やその混在比率によらずにアクセス帯域が大きいコネクションの利用可能帯域はアクセス帯域より小さくなる．したがって，このような置き換えで近似しても，大きいアクセスリンク帯域を持つコネクションのファイル転送時間が増大し始めるようなボトルネックリンク使用率を精度よく近似できる(詳細は文献[8]参照)．

4.3 アクセス帯域以外の要因を考慮したモデル

節4.1ではアクセスリンク帯域ネックのためにボトルネックリンク帯域を使いきれない場合を扱っていたが，往復遅延ネックの場合，すなわち往復伝搬遅延が大きく，$W_M/D<C$ となる場合にも同様にモデル化できると考えられる．すなわち伝播遅延ネックの場合の単一コネクション時の上限スループットは W_M/D であるから，これを仮想的にアクセス帯域 r と考えれば，処理能力 r，サーバ数 C/r の複数サーバPS待ち行列でモデル化できる．文献[9]では，このモデル化を用いた帯域管理法が，C/r の測定方法も含めて提案されている．

また節4.2においても，ボトルネックリンクが輻輳していなければ各コネクションは常にアクセスリンク帯域の速度でデータ送信が可能であるとしてモデル化していた．しかし，TCPコネクションの実際の送信速度は，往復伝播遅延時間やウインドウサイズ，図1で着目するボトルネックリンク以外での輻輳に起因する遅延・パケット損失によっても制限され，ボトルネックリンクが輻輳していなくてもアクセスリンク帯域を使い切れない場合がある．さらに，これらの要因は個々のコネクションによって異なり，各コネクションの送信速度も様々となる．

そこで文献[8]では，着目するコネクションに対し，ボトルネックリンク使用率が十分小さいときのTCPスループットを実測を通じて求めておき，それを上限スループットとして節4.2のアクセスリンク帯域 r' に代入する（すなわち，上限スループットを仮想的にアクセス帯域とみなす）ことによって，ボトルネックリンク使用率が増加したときのファイル転送時間を推定する方法を提案している．図3に，あるボトルネックリンクにおける実測評価結果を示す．この図は，当該リンクを経由してTCPでファイル転送を実行し，リンク非混雑時のスループットを r' として，上記提案式で混雑時（ボトルネックリンク使用率上昇時）の平均転送時間を推定したものと，実際の転送時間を比較したものである．

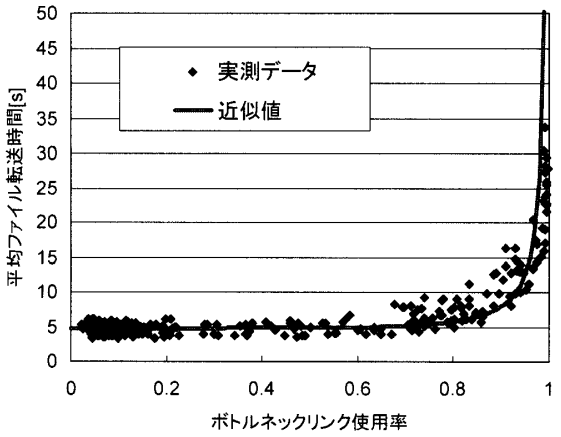

図3 異なる帯域制限を持つコネクション混在時の特定コネクション平均転送時間近似

5. おわりに

本稿では，プロセッサシェアリング待ち行列を用いて，TCPコネクション間でリンク帯域が公平に配分されることをモデル化してTCP性能を評価する方法について説明した．また，対象とするネットワーク環境に応じて，モデルを拡張する試みについてもいくつか紹介した．

最後に他の拡張に関しても若干触れておく．本稿では，他のコネクションとの帯域競合が発生するリンクは一つとしたが，実際には複数のリンクで他のコネクションと帯域競合が発生する場合も考えられる．この場合，あるリンクの利用可能帯域は，そのリンクの同時接続コネクション数のみならず，それらのコネクションが経由する他のリンクにおける帯域配分結果にも依存する．このような状況におけるコネクション間の帯域配分については，主に公平性の観点から広く研究されており，様々な種類の公平性が提示されている[10]．節4.2に述べたように，このような場合のTCPコネクションの性能を求めることは困難であるが，近年性能近似法に関する研究も報告されている[11]．

また，本稿ではボトルネックはデータ送信方向にのみ存在すると仮定したが，実際にはAck返信方向の経路が輻輳することもありうる．P2Pなどの双方向通信の普及に伴い，ネットワーク設計のためには，このようなAck返信方向での輻輳の影響を考慮したモデル化も必要になってくると思われる．

参考文献

[1] M. Mathis, J. Semke, J. Mahdavi and T. Ott: "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", Computer Communications Review, vol. 27, no. 3, 1997.

[2] J. Padhye, V. Firoiu, D. Towsley, J. Kurose: "Modeling TCP Throughput: A Simple Model and its Empirical Validation", Proc. ACM SIGCOMM'98, 1998.

[3] 長谷川剛, 村田正幸, 宮原秀夫: "TCPの公平性と安定性に関する一検討", 電子情報通信学会論文誌 B-I, vol. J 82-B-1, no. 1 pp. 1-9, 1999.

[4] L. Massoulie and J. Roberts: "Bandwidth sharing and admission control for elastic traffic", Telecommunication Systems, vol. 15, no. 1/2, 2000.

[5] D. P. Heyman, T. V. Lakshman, and A. L. Neidhardt: "A new method for analysing feedback-based protocols with applications to engineering Web traffic over the Internet", Proc. ACM SIGMETRICS 1997, pp. 24-38, 1997.

[6] S. B. Fredj, T. Bonald, A. Proutiere, G. Regnie, and J. W. Roberts: "Statistical bandwidth sharing: a study of congestion at flow level", Proc. ACM SIGCOMM 2001, 2001.

[7] D. Bertsekas and R. Gallager: Data networks, 2nd ed., Prentice-Hall, Upper Saddle River, 1992.

[8] 川原亮一, 石橋圭介, 森達哉, 小沢利久, 住田修一, 阿部威郎: "異速度TCPフロー集約リンクにおけるTCP品質推定法と帯域設計管理法", 日本OR学会春季研究発表会, 2004.

[9] 川原亮一, 石橋圭介, 朝香卓也: "フロー統計情報を用いた帯域設計管理法", 電子情報通信学会信学技報 NS 2003-87, 2003.

[10] J. Mo and J. Walrand: "Fair end-to-end window-based congestion control", IEEE/ACM Trans. on Networking, vol. 8 no. 5, pp. 556-567, 2000.

[11] G. Fayolle, A. L. Fortelle, J.-M. Lasgouttes, L. Massoulie, and J. Roberts: "Best-effort networks: modeling and performance analysis via large networks asymptotics", Proc. IEEE INFOCOM 2001, 2001.