

特別研究報告

題目

ルータにおけるバッファサイズが
TCP性能に与える影響のシミュレーション評価

指導教員

中野 博隆 教授

報告者

富岡 健史

平成 18 年 2 月 27 日

大阪大学 基礎工学部 情報科学科

ルータにおけるバッファサイズが
TCP 性能に与える影響のシミュレーション評価

富岡 健史

内容梗概

現在、インターネットルータのバッファ容量の決定のためには、帯域遅延積を指標とする方法が広く利用されている。これに対し、ルータを通過するフローが多く存在する場合、ネットワークリンクの利用率を維持するには帯域遅延積をフロー数の平方根で除算したサイズで十分であるという主張が提起されている。しかしこの主張には、リンク利用率以外の視点からの評価が行われていない、評価を行う際のパラメータが限られている等の問題が存在する。そこで本報告では、ns-2 を用いたシミュレーションにより、ルータのバッファ容量が TCP の性能に与える影響を考察する。その結果、フロー数が十分に大きければ、バッファ容量を減少してもリンク利用率を高く維持できるが、帯域遅延積を用いた場合と比較すると、パケット廃棄率に関して性能が劣化していること、50 - 100Kbyte 程度の転送データサイズを境としてデータ転送遅延時間が増加し、転送データサイズが小さいフローが多数を占めるネットワークや、エンド端末間の伝播遅延時間が小さいネットワーク以外においては、バッファ容量を減少することによる悪影響が現れることが明らかになった。

主な用語

ルータ、バッファ容量、TCP (Transmission Control Protocol)、リンク帯域、伝播遅延時間

目次

1	はじめに	5
2	ルータのバッファ容量の決定方法	8
2.1	帯域遅延積によるバッファ容量の決定	8
2.2	コネクション数に応じたバッファ容量の削減手法	11
3	シミュレーション環境	13
3.1	ネットワークとトラヒックモデル	13
3.2	シミュレーション結果の評価方法	15
4	シミュレーション結果と考察	16
4.1	基本性能	16
4.2	ネットワーク環境の変化が与える影響	18
4.2.1	ネットワークトラヒック量	18
4.2.2	アクセスリンク帯域	21
4.2.3	ボトルネックリンク帯域	24
4.2.4	ボトルネックリンクの伝播遅延時間	27
5	おわりに	29
	謝辞	30
	参考文献	31

目次

1	一本の TCP フローが流れている場合の、輻輳ウィンドウサイズ及びバッファ内パケット数の変化	10
2	シミュレーションで用いるネットワークモデル	14
3	バッファ容量がネットワークのリンク利用率とパケット廃棄率に与える影響	17
4	ボトルネックリンクの伝播遅延時間が小さい場合 ($D=20$ ms) の、ホスト数に与える影響	19
5	ボトルネックリンクの伝播遅延時間が大きい場合 ($D=90$ ms) の、ホスト数に与える影響	20
6	アクセスリンク帯域がパケット廃棄率に与える影響	22
7	アクセスリンク帯域がデータ転送遅延時間に与える影響	23
8	ボトルネックリンク帯域がネットワーク性能に与える影響	25
9	ボトルネックリンク帯域がデータ転送遅延時間に与える影響	26
10	ボトルネックリンクの伝播遅延時間が与える影響	28

表目次

1	[1]に基づく Web アクセスモデル (SURGE モデル)	14
---	---	----

1 はじめに

TCP (Transmission Control Protocol) [2] は現在インターネットで広く利用されており、TCP トラヒックは現在のインターネットトラヒックの大部分を占める [3]。他に利用されるプロトコルとしては UDP (User Datagram Protocol) [4] があげられるが、TCP と比較すればトラヒックに占める割合は非常に小さい。そのため、ネットワーク上の TCP トラヒックの性能評価を行うことは非常に重要である。TCP によるデータ転送の性能は、パケット廃棄率やラウンドトリップ時間 (RTT) によって大きく影響を受ける [5, 6]。現在のインターネットルータの多くは First-in First-out (FIFO) 規律に基づくバッファを持つが、FIFO 規律に基づくバッファを用いている場合において、ネットワークのパケット廃棄率やデータ転送遅延時間を決定する大きな要素の一つとして、ルータのバッファ容量が挙げられる。リンク帯域を越えるレートのパケット送信がそのリンクに発生した場合、パケットは一時的にそのリンクの出力バッファ内に到着順に蓄積される。バッファが一杯のときに新たなパケットが到着すると、そのパケットは廃棄される。そのため、一般にバッファサイズを大容量にすることで、パケット廃棄が発生する確率を下げることができる。しかし、バッファ内に大量のデータが蓄積されるようになれば、パケットがバッファに到着してから、処理を受けてリンクへ出力されるまでの時間、すなわちキューイング遅延が増大する。[7] によると、リンク利用率を高い水準で維持しつつ、キューイング遅延を最小にするためのボトルネックリンクのバッファ容量は、リンク帯域 C とそのルータを通過する TCP コネクション (フロー) の平均 RTT (以下 \overline{RTT} と表記する) を用いて $B_n = C \times \overline{RTT}$ であるとされている。これは [8] などでも述べられている。このバッファ容量の決定方法は帯域遅延積として知られており、現在多くのルータがこの方法に従った容量のバッファを装備している。[9] にも、ルータは一般に帯域遅延積で決定される容量のバッファを持つ、と記されている。

しかし [10] によれば、この方法に基づく容量のバッファを持つルータを構築することは困難になっているという。現在のバックボーンネットワークでは 2.5 Gbps や 10 Gbps といった帯域幅を持つリンクに、10,000 を超えるフローが流れている [11]。10 Gbps の帯域を持つリンクに流れるフローの平均 RTT が 250 ms であれば、必要とされるバッファ容量は $10 \text{ Gbps} \times 250 \text{ ms} = 2.5 \text{ Gbits}$ となる。ルータのバッファに利用されるメモリデバイスは SRAM と DRAM の二種類が存在するが、現在商用に用いられている SRAM チップの最大のサイズは 36 Mbits であり、必要なバッファ容量を確保するためには多数の SRAM チップを並列的に用いて構築する必要がある。たとえば 2.5 Gbits のバッファを構築するためには、およそ 70 個の SRAM チップが必要となる。しかし SRAM は DRAM に比べて非常に高価であるため、SRAM を用いてバッファを構築することは経済的に困難である。また、必要とされる設置面積が増大すること、発熱量が大きくなることなど物理的にも多くの問題があ

る。一方 DRAM は SRAM よりも安価であり、最大 1 Gbits のサイズを持つが、アクセス時間が 50 ns と大きいということ、使用するにつれて 7 %/年の割合でアクセス時間が低下することなどから、やはり大容量のバッファを構成することは難しい [12]。これらのことから、10 Gbps を超えるリンクにおいて、帯域遅延積に相当する容量のバッファを装備することは難しい、と [10] は述べている。

さらに [10] ではその解決方法として、リンクを通過するフロー数が十分に多い場合（主張では 500 程度以上）には、バッファ容量は帯域遅延積に対しリンクを通過するフロー数 N の平方根で除算した値、すなわち $B_s = \frac{C \times RTT}{\sqrt{N}}$ で十分であるということを示している。[10] で示されている結果によると、この方法でバッファ容量を設定した場合にもリンク利用率はほとんど低下しない。また、[13] では、アクセスリンクの帯域が非常に小さい場合（1 Gbps のボトルネックリンク帯域に対して 10 Mbps 程度のアクセスリンク帯域を持つ場合）や、ペーシング TCP（TCP のデータパケットがバースト的に送出されるのを防止した手法）を利用した場合には、さらに小さい数十パケットのバッファ容量で十分であるとされている。

しかし、[10, 13] はともに、性能指標としてボトルネックリンクの利用率のみに着目している。リンク利用率を保つことはネットワーク管理者にとっては重要であるが、ネットワークを利用するユーザにとって重要な性能指標である TCP の性能（フローの packets 廃棄率やデータ転送遅延時間）に関する評価も必要であると考えられる。[10, 13] におけるシミュレーションおよび実験の結果はリンク利用率が維持されていることを示すものがほとんどであり、用いられているネットワークに関するパラメータも限定されている。また [10] では、転送データサイズが 14 パケットの場合についてのみ、データ転送遅延時間に関するシミュレーション結果が示されており、バッファ容量を減少させることによってデータ転送遅延時間が小さくなるとされている。しかし、パケット廃棄率については、TCP の性質から数 RTT に一度しか廃棄が発生しないため影響は少ないのではないか、という見解のみが述べられている。また [13] で示された結果から、アクセスリンク帯域が小さい場合、あるいはペーシング TCP を用いる場合においては、TCP がデータを送信する際のバースト性が軽減されるため、パケット廃棄率およびデータ転送遅延時間が減少すると考えられるが、アクセスネットワークのリンク帯域は近年急速に増大している上に、現在ペーシング TCP はほとんど普及しておらず、今後も急速に普及することは期待できない。

そこで本報告では、TCP-Reno が利用される環境の下で、ルータのバッファ容量がネットワークのリンク利用率やパケット廃棄率、および個々のフローの packets 廃棄率やデータ転送時間に与える影響を、計算機上のシミュレーションによって評価する。評価に際してはボトルネックリンクの伝播遅延時間やリンク帯域、データ転送を行うホスト数などをパラメータとしてシミュレーションを行い、[10] で提案された手法でバッファ容量を決定した場合と、帯域遅延積を用いて決定した場合を、パケット廃棄率、データ転送遅延時間を性能指標とし

て比較することで、バッファ容量の変化による TCP の性能の変化と、TCP 性能から見た場合にどのバッファ容量が望ましいのか明らかにする。

本報告の構成は以下の通りである。2 章で本報告で対象とする 2 種類のバッファ容量の決定指標について詳細を示す。次に 3 章でシミュレーションに用いたネットワークの構成、各種パラメータ、および性能指標について説明を行う。4 章ではシミュレーション結果を示し、考察を行う。最後に 5 章でまとめと今後の課題を示す。

2 ルータのバッファ容量の決定方法

2.1 帯域遅延積によるバッファ容量の決定

帯域遅延積によるバッファ容量の決定は [7] で提唱され、現在広く用いられている。本報告では以降この方法を normal 指標と称する。

Reno [8] や NewReno [14] などインターネット上で一般的に使われる TCP では、AIMD (Additive-increase Multiple-decrease) と呼ばれる方法 [15] で輻輳ウィンドウサイズの制御を行っている。図 1 は一本の TCP フローが輻輳回避フェーズでデータ転送を行っている場合の、TCP フローの輻輳ウィンドウサイズとバッファ内パケット数の変化を示している。ここで B_{max} はバッファ容量をあらわす。時刻 t_1 から TCP のパケット送信レートがリンク帯域を超え、リンク帯域を越えて到着したパケットがバッファ内に蓄積され始める。TCP は ACK パケットを受け取るたびに輻輳ウィンドウサイズを増加させるため、最終的にはバッファあふれが発生しパケットが廃棄される (時刻 t_2)。パケットが廃棄されてから約 1RTT 後には、送信側 TCP が重複 ACK を受信することによってパケット廃棄を検知し、Fast Retransmit 機構 [6] によりそのパケットを再送すると同時に、輻輳ウィンドウサイズを廃棄が発生したときの値 W_{max} から $W_{max}/2$ に減少させる (時刻 t_3)。この結果、送信側 TCP から送出されるパケット数が減少し、ルータへのパケット到着レートがリンク帯域を下回るため、バッファ内に蓄積されていたパケット数は減少する (時刻 $t_3 - t_5$)。その後、送信側 TCP は輻輳ウィンドウサイズを再び増加させるため、時刻 t_5 からバッファ内のパケット数は再び増加に転じる。

このとき、バッファ内のパケット数が時刻 t_5 に達する前に 0 になると、ルータへのパケットの流入量がリンク帯域を下回りかつバッファ内にパケットが存在しない状態、すなわち送信すべきデータが存在するにもかかわらずリンク帯域が 100% 利用されていない状態が生じる。

リンク帯域 C のボトルネックリンクにおいてルータのバッファが空にならない場合、送信ノードへの ACK パケット到着レートは C で表すことができる。一方、輻輳ウィンドウサイズ W_{max} から $W_{max}/2$ に減少した場合、送信側 TCP は $W_{max}/2$ 個の ACK パケットが帰ってくるまで、新しいパケットを送信することができない。送信ノードに $W_{max}/2$ 個の ACK パケットが到着するまでにかかる時間は $(W_{max}/2)/C$ で表すことができる。また、バッファ容量 B_{max} のバッファが空になるまでには B_{max}/C の時間がかかる。すなわち、バッファが空になるための条件は、

$$B_{max}/C < (W_{max}/2)/C \quad (1)$$

と表すことができる。逆に $B_{max}/C \geq (W_{max}/2)/C$ であればバッファが空になる前に送信

側からのパケット送信が再開されるため、リンク利用率を維持することができる。すなわち、バッファが空になる前に新しいパケットが送信側 TCP から到着するためには、次式で表されるバッファ容量が必要となる。

$$B_{max} \geq W_{max}/2 \quad (2)$$

また、リンク上に存在するデータの総量は \overline{RTT} と C を用いて $C \times \overline{RTT}$ として表すことができるが、輻輳ウィンドウサイズの値が、半減した後も $C \times \overline{RTT}$ 以上であれば、リンクに常にデータが流れるためにリンク利用率を維持することができる。すなわち、リンク利用率を維持するためには、

$$W_{max}/2 = C \times \overline{RTT} \quad (3)$$

であればよい。(2)(3) 式から

$$B_{max} \geq W_{max}/2 = C \times \overline{RTT} \quad (4)$$

となる。すなわち、1本の TCP フローが存在する場合にリンク利用率を維持するためには、 $C \times \overline{RTT}$ のサイズのバッファが必要であるといえる。

また、TCP フローが複数本存在する場合にも、それらが同期して動作する、すなわち、バッファあふれが発生したときにすべてのフローのパケットが等しく廃棄される場合には、必要となるバッファ容量は変化しない。これは、TCP フローが N 本存在する場合には、個々のフローは $1/N$ の帯域を持つリンクに1本のフローを流したときと同様の挙動を示す [16-19] ため、 $C/N \times \overline{RTT}$ のバッファが N 本分必要と見なすことができ、全体として必要となるバッファ容量はやはり、

$$(C/N \times \overline{RTT}) \times N = C \times \overline{RTT} \quad (5)$$

と表すことができる。

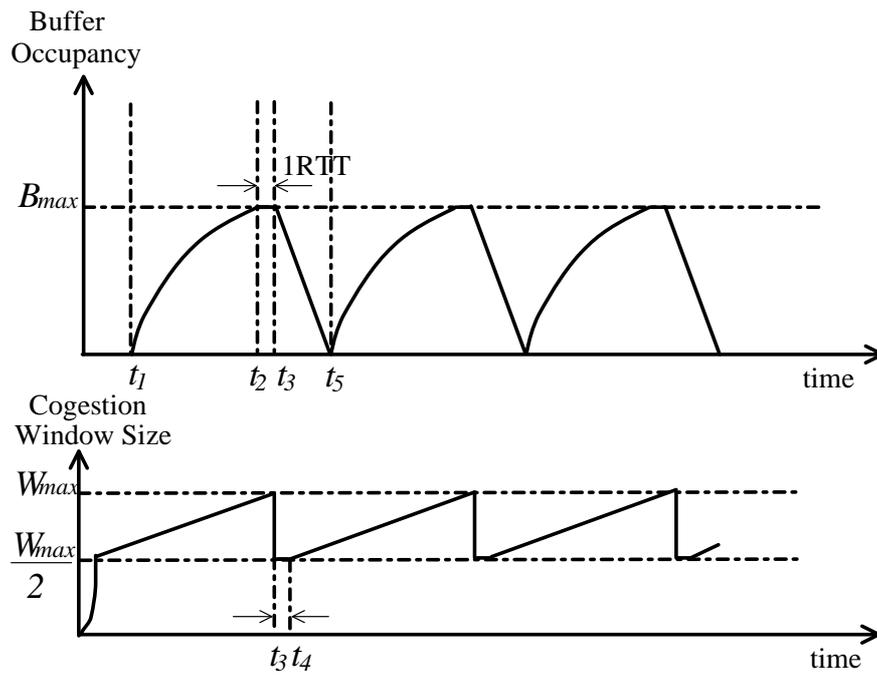


図 1: 一本の TCP フローが流れている場合の、輻輳ウィンドウサイズ及びバッファ内パケット数の変化

2.2 コネクション数に応じたバッファ容量の削減手法

この指標は、帯域遅延積をリンクを通過するフロー数の平方根で除算したものをバッファ容量とするものである。本報告ではこの指標を sqrtN 指標と称する。以下では、[10] で示された sqrtN 指標の根拠の概要を示す。詳細については [10] を参照されたい。

TCP の輻輳ウィンドウサイズは通常、同期して変動することが知られている [16–19]。すなわち、あるフローの輻輳ウィンドウサイズが半減するときには、他のフローの輻輳ウィンドウサイズも半減する [16]。これは、ルータのバッファあふれによってパケット廃棄が発生する際に、通過するフローのパケットが等しく廃棄されることによる。しかし、各フローの RTT がそれぞれ異なるような場合にはこの性質が失われ、輻輳ウィンドウサイズの変動は非同期となる [20]。これは、バッファあふれによるパケット廃棄が全フローに同期して発生しないことを意味し、現実のネットワーク上でも見られる [11, 21]。[10] によると、RTT がそれぞれ異なっていなくとも、ボトルネックリンクを通過するフロー数がおよそ 100 以下の場合には TCP の輻輳ウィンドウサイズは同期して変動するが、フロー数が 500 を超えるような場合は、輻輳ウィンドウサイズがの同期性が失われる。このとき、ルータを通過する TCP コネクションの輻輳ウィンドウサイズの分布はほぼ正規分布にしたがう。

N 本のフローが通過するルータの時刻 t におけるキュー長 (バッファ内のパケット数) $Q(t)$ は、その時の各フローの輻輳ウィンドウサイズ $W_i(t)$ を用いて

$$Q(t) = \max(0, \sum_{i=1}^N W_i(t) - (\overline{RTT} \times C))$$

と表すことができる。よって、 $\sum_{i=1}^N \overline{W_i(t)} = \overline{W}$ を用いて、ルータにおける平均のキュー長 \overline{Q} は

$$\overline{Q} = \max(0, \overline{W} - (\overline{RTT} \times C)) \quad (6)$$

となる。すなわち $\overline{Q} > 0$ のとき、各フローの輻輳ウィンドウサイズの平均値 $\overline{W_i}$ は (6) 式を用いて

$$\overline{W_i} = \overline{W}/N = \frac{\overline{RTT} \times C + \overline{Q}}{N} \leq \frac{\overline{RTT} \times C + B_{max}}{N} \quad (7)$$

と求めることができる [10]。また TCP の輻輳ウィンドウサイズの平均値 $\overline{W_i}$ に対する標準偏差 σ_{w_i} は、TCP フローのウィンドウサイズの変動をモデル化することによって、

$$\sigma_{w_i} = \frac{1}{3\sqrt{3}} \overline{W_i} \quad (8)$$

となる。正規分布の性質より、フロー数が十分に多い場合は、輻輳ウィンドウサイズの合計

値の標準偏差 σ_w が

$$\sigma_w \leq \sqrt{n}\sigma_{w_i} \quad (9)$$

となることから、ルータにおけるキュー長 Q の標準偏差 σ_Q は、(7)-(9) 式より

$$\sigma_Q = \sigma_w \leq \frac{1}{3\sqrt{3}} \frac{\overline{RTT} \times C + Q}{\sqrt{n}} \leq \frac{\overline{RTT} \times C + B_{max}}{\sqrt{n}} \quad (10)$$

と表すことができる。これらのことから、リンク利用率 $Util$ について次式を導くことができる。

$$Util \geq erf \left(\frac{3\sqrt{3}}{2\sqrt{2}} \frac{B_{max}}{\frac{\overline{RTT} \times C + B_{max}}{\sqrt{n}}} \right) \quad (11)$$

$B_{max} = \frac{\overline{RTT} \times C}{\sqrt{n}}$ であれば $Util \geq erf\left(\frac{3\sqrt{3}}{2\sqrt{2}}\right) \simeq 0.9899$ となるため、フロー数が十分に多い場合は $B_s = \frac{\overline{RTT} \times C}{\sqrt{n}}$ で十分である。

また [10] では、スロースタートフェーズ中にデータ転送が終わるような転送データサイズの小さいフロー (short-lived フロー) が存在する場合においても、必要となるバッファ容量は short-lived フローにはほとんど影響を受けず、 $B_s = \frac{\overline{RTT} \times C}{\sqrt{n}}$ で導かれるバッファ容量より、とされている。

3 シミュレーション環境

3.1 ネットワークとトラフィックモデル

シミュレーションでは図2のネットワークモデルを用いる。モデルは送信側ホスト $S_1 \sim S_n$ 、受信側ホスト $R_1 \sim R_n$ と両者をつなぐアクセスリンク、ボトルネックリンク、ルータで構成される。ボトルネックリンクの伝播遅延時間および帯域をそれぞれ D [ms] および C [Mbps] とし、アクセスリンクの帯域はボトルネックリンクと等しいものとし、変化させる場合にはそれを明記する。シミュレーションにおいては、ホスト数 N を変化させ、ボトルネックリンクを通過するフロー数を変動させる。また、 D 、 C についても変化させ、各条件の下でバッファ容量がTCPの性能に与える影響を評価する。各ホストから送出されるトラフィックには、各ホストが無限大の転送データサイズをもつP2Pトラフィック、および[1]で述べられているSURGEモデル(表1)にしたがって、各ホストの転送データサイズや転送間隔などが決定されるWebトラフィックを用いる。以降、明記していない場合はWebトラフィックを使用する。シミュレーションにはns-2 [22]を用いる。

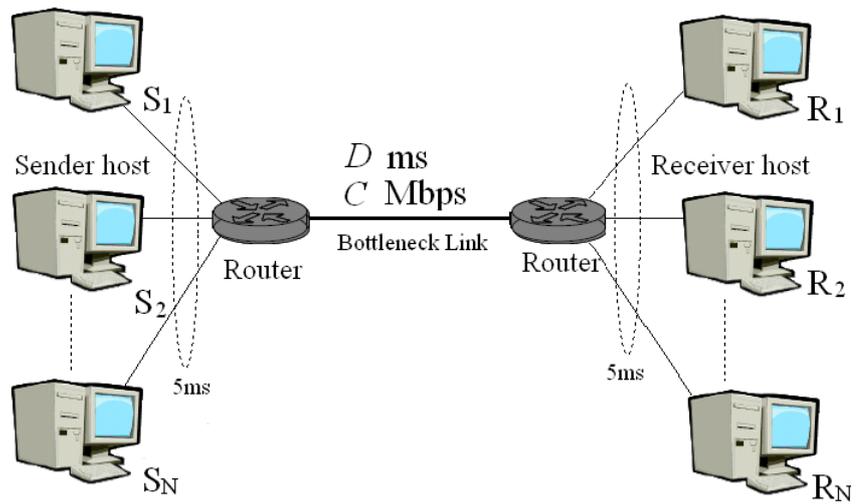


図 2: シミュレーションで用いるネットワークモデル

表 1: [1] に基づく Web アクセスモデル (SURGE モデル)

要素	生成関数	パラメータ
サイズ - Body	$p(x) = \frac{e^{-(\ln x - \mu)^2 / 2\sigma^2}}{x\sigma\sqrt{2\pi}}$	$\mu = 9.357; \sigma = 1.318$
サイズ - Tail	$p(x) = \alpha k^\alpha x^{-\alpha+1}$	$k = 133K; \alpha = 1.1$
生成間隔	$p(x) = \alpha k^\alpha x^{-\alpha+1}$	$k = 1; \alpha = 1.5$

3.2 シミュレーション結果の評価方法

ボトルネックルータのリンクにおいて、通過するパケットの挙動を観測し、単位時間当たりには通過したパケット数からリンク利用率を求める。また、ルータに到着したパケット数とバッファあふれによって廃棄されたパケット数から、パケット廃棄率を導出する。Webトラフィックを用いたシミュレーションの場合には、以下に示す各フローの転送に関する性能指標を用いる。データ転送遅延時間は、各フローが転送するデータの最初のパケットの送信が開始されてから、最後のパケットに対する ACK を受信するまでの時間とし、各ホストのパケット送受信のログより求める。また、転送したデータサイズを記録し、データ転送が終了するまでに送信したパケット数との比較から、各転送において廃棄されたパケット数を求め、各データサイズにおけるパケット廃棄率を導出する。このようにして得られたリンク利用率、パケット廃棄率、各サイズにおけるパケット廃棄率、およびデータ転送遅延時間を性能評価指数として用いる。

4 シミュレーション結果と考察

4.1 基本性能

図3は $C=100$ Mbps、 $D=90$ ms のネットワークに P2P トラフィックおよび Web トラフィックをそれぞれ与えた環境において、ホスト数を変化させた場合のリンク利用率とパケット廃棄率の変化を示したものである。図3(a)から、normal 指標でバッファ容量を決定した場合は、ホスト数が少ない場合においても高いリンク利用率を示しているが、sqrtN 指標を用いた場合は、ホスト数が小さい場合において P2P トラフィックの場合（図3(a)）で最大 20% 程度、Web トラフィックの場合（図3(b)）においても最大約 10% 程度のリンク利用率の低下が生じている。特に Web トラフィックの場合（図3(b)）において、リンク利用率が 60% 程度の場合、すなわちネットワークにかかっている負荷が高くない場合からリンク利用率の低下が生じていることから、TCP が同期して動作している場合に sqrtN 指標をバッファ容量の決定に利用することはネットワーク性能に悪影響をおよぼすといえる。

一方、パケット廃棄率に関しては、一貫して sqrtN 指標を用いた場合のほうが高くなっている。特に、図3(b)から、normal 指標ではホスト数が 700 程度より小さい場合にはパケット廃棄が発生していないのに対し、sqrtN 指標ではホスト数に関わらず廃棄が発生していることが確認できる。これは、sqrtN 指標を用いるとバッファ容量が減少するため、パケットがバースト的に流入した時にバッファで吸収しきれずにバッファあふれが発生し、パケットが廃棄されてしまうためである。

しかし、トラフィックの種類に関わらず、ホスト数が多い場合にはリンク利用率はほぼ 100% であり、normal 指標を用いた場合と差がなくなっていることから、[10]における主張が正しいことが裏づけられる。

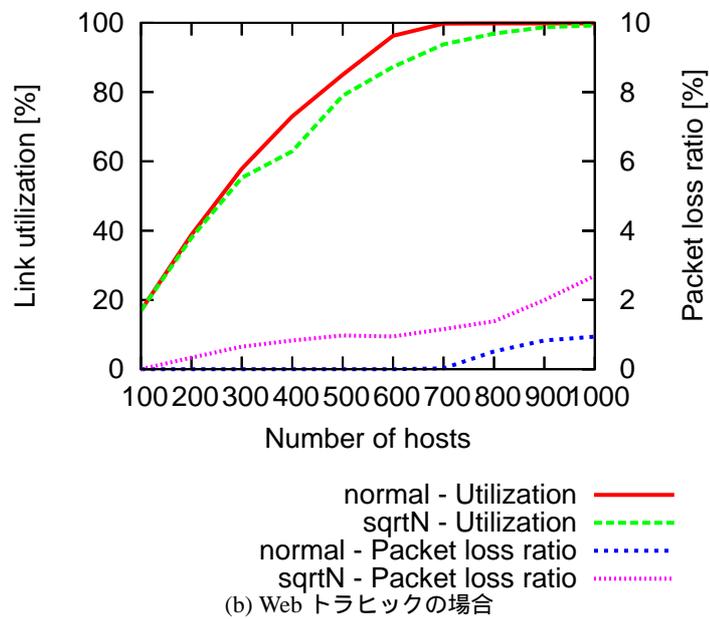
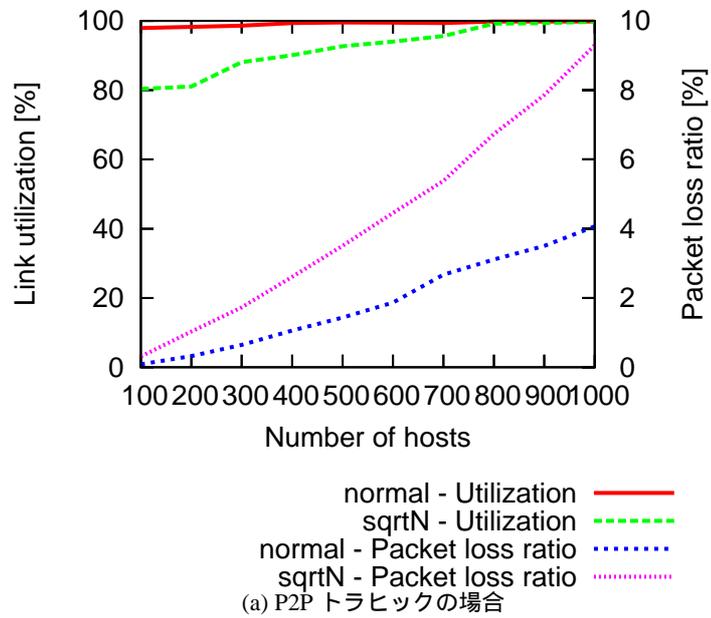


図 3: バッファ容量がネットワークのリンク利用率とパケット廃棄率に与える影響

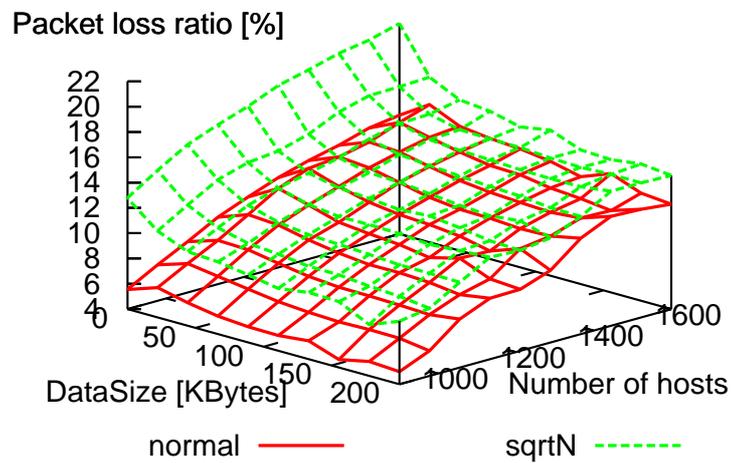
4.2 ネットワーク環境の変化が与える影響

本節では、ネットワーク環境が変化した場合に、ルータのバッファ容量の決定に normal 指標および sqrtN 指標のどちらを適用すべきかについて議論する。これにより、将来のネットワークにおけるルータのバッファ容量に関する知見を得ることを目的とする。

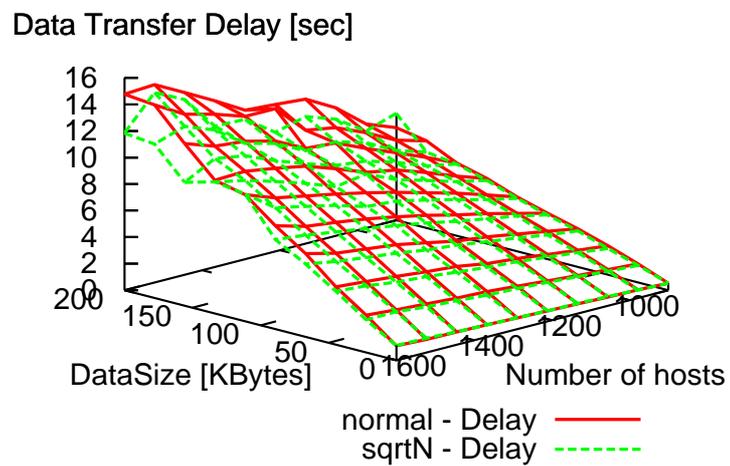
4.2.1 ネットワークトラヒック量

図 4 および図 5 は、それぞれ $C=100$ Mbps、 $D=20$ ms および $C=100$ Mbps、 $D=90$ ms の場合においてホスト数を変化させたときの、転送データサイズに対するパケット廃棄率およびデータ転送遅延時間の変化を示したものである。図 4(a) および図 5(a) から、sqrtN 指標を用いた場合は、転送データサイズに関係なくパケット廃棄率が normal 指標にくらべて高いことがわかる。しかし、ホスト数が増加するにつれ、両指標間のパケット廃棄率の差が小さくなっているのに対して、図 5(a) においては、ホスト数が増加するにつれて量指標間のパケット廃棄率の差が拡大している。また、図 4(a) においては、ホスト数にかかわらず、小さい転送データサイズの場合にパケット廃棄が多く発生している。一方、データ転送遅延時間（図 4(b)、図 5(b)）においては、ホスト数の増加とデータ転送遅延時間の増加はほぼ比例している。これは normal 指標、sqrtN 指標のいずれにおいても同じであり、両指標間のホスト数の変化による差異は小さいといえる。図 4(b) ではホスト数が増加するにしたがい、データ転送遅延時間の差が大きくなっているが、これはパケット廃棄率の差が増大することに起因すると考えられる。それに対して図 10(b) ではホスト数の変化によるパケット廃棄率の差が変化していないため、データ転送遅延時間の差にも大きな変化は見られない。

これらのことから、ホスト数が増加しネットワークトラヒック量が増加する場合においては、トラヒック量そのものではなく、ボトルネックリンク、アクセスリンクの帯域や、ボトルネックリンクの伝播遅延時間などの影響を考慮し、normal 指標を用いるか sqrtN 指標を用いるかを決定すればよいと考えられる。

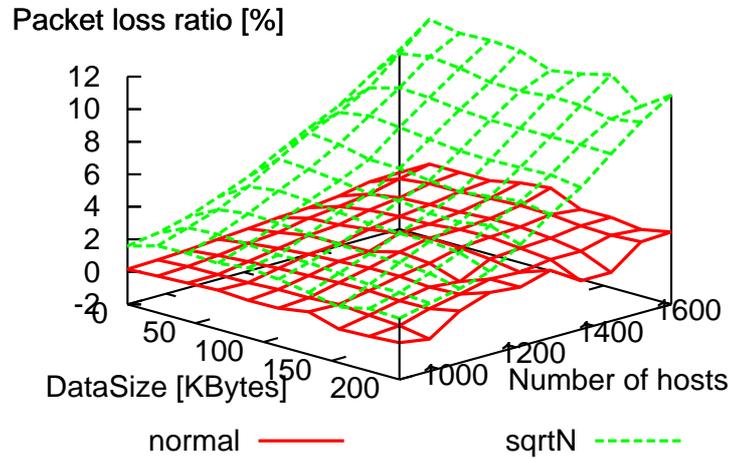


(a) パケット廃棄率

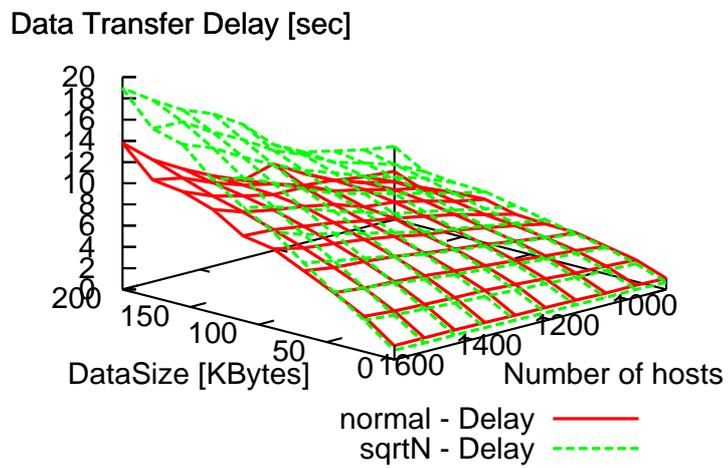


(b) データ転送遅延時間

図 4: ボトルネックリンクの伝播遅延時間が小さい場合 ($D=20$ ms) の、ホスト数が与える影響



(a) パケット廃棄率



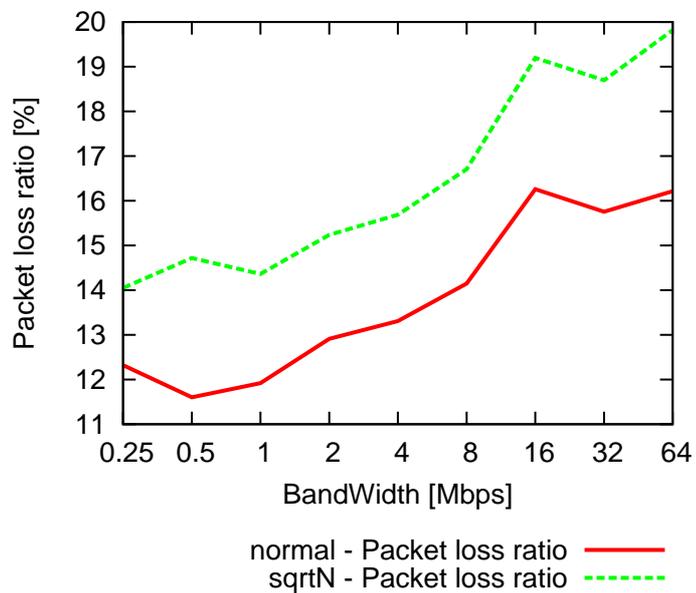
(b) データ転送遅延時間

図 5: ボトルネックリンクの伝播遅延時間が大きい場合 ($D=90$ ms) の、ホスト数が与える影響

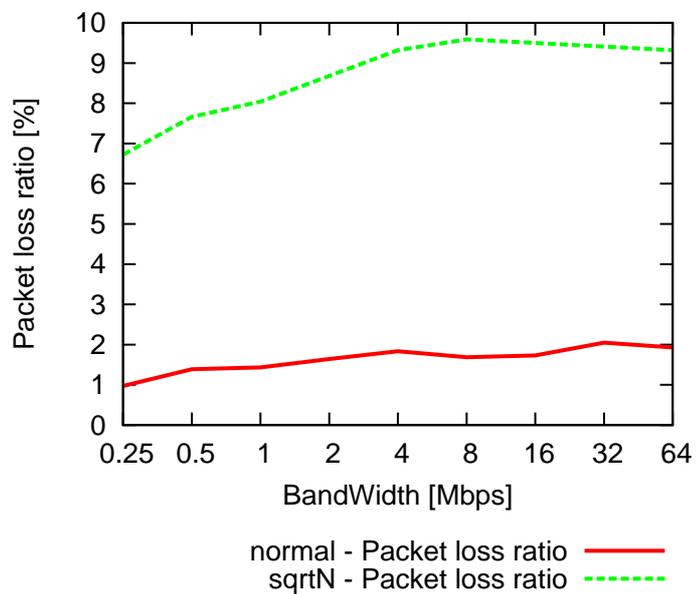
4.2.2 アクセスリンク帯域

図 6 および図 7 は、それぞれ $C=100$ Mbps、 $D=20$ ms、 $N=1500$ および $C=100$ Mbps、 $D=90$ ms、 $N=1500$ の場合においてアクセスリンクの帯域を変化させたときの、転送データサイズに対するパケット廃棄率およびデータ転送遅延時間の変化を示したものである。図 6(a) を見ると、アクセスリンク帯域が変化しても、normal 指標を利用した場合と sqrtN 指標を利用した場合のパケット廃棄率の差が変化していないが、図 6(b) からは、アクセスリンク帯域が 8 Mbps 程度まではアクセスリンク帯域の増加につれて、両者間の差が開いていることがわかる。これは、アクセスリンク帯域がボトルネックリンク帯域に対して十分に細いときには、通過する TCP トラヒックのバースト性が低下することが原因である [13] と考えられる。そのため、図 6(b) でアクセスリンク帯域が増加するにつれ、バースト性が見られるようになっていく。また、図 6(a) で両者間の差が変化しない理由は、ボトルネックリンクの伝播遅延時間が短いために、アクセスリンク帯域が小さい場合にも TCP トラヒックのバースト性が失われないためだと考えられる。図 7 に示すデータ転送遅延時間の変化もパケット廃棄率の変化を反映しており、図 7(a) では両者間データ転送遅延時間の差はほとんど変化していないが、図 7(b) ではアクセスリンク帯域が大きくなるに従って、データ転送遅延時間の差が大きくなっている。

以上のことから、アクセスリンク帯域が増加する場合には sqrtN 指標を用いることでネットワークを通過する TCP フローに対して性能の低下が発生する可能性があるといえる。しかし、アクセスリンク帯域の増加がもたらす影響の大きさはボトルネックの伝播遅延時間によって変化すること、ボトルネックリンク帯域が大きくなれば相対的にはアクセスリンクの大きさは低下することから、ボトルネックリンクの伝播遅延時間や帯域がより重要な判断基準になると考えられる。

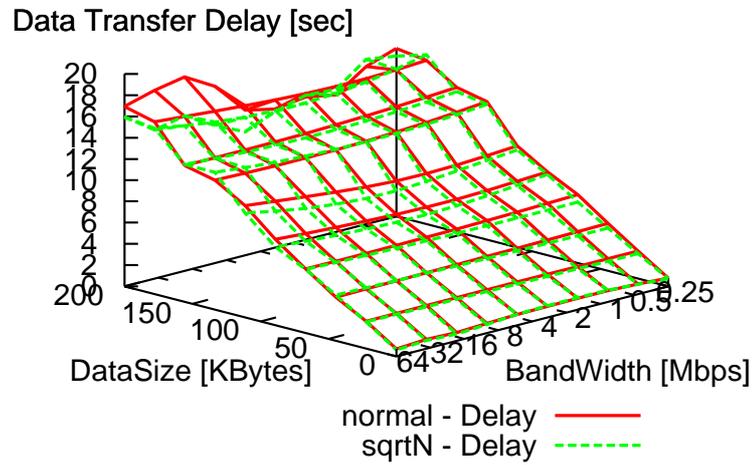


(a) ボトルネックリンクの伝播遅延時間が小さい場合 ($D=20$ ms)

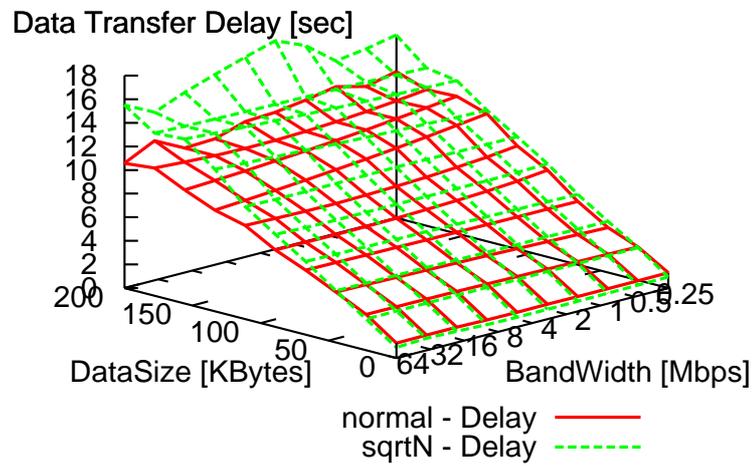


(b) ボトルネックリンクの伝播遅延時間が大きい場合 ($D=90$ ms)

図 6: アクセスリンク帯域がパケット廃棄率に与える影響



(a) ボトルネックリンクの伝播遅延時間が小さい場合 ($D=20$ ms)



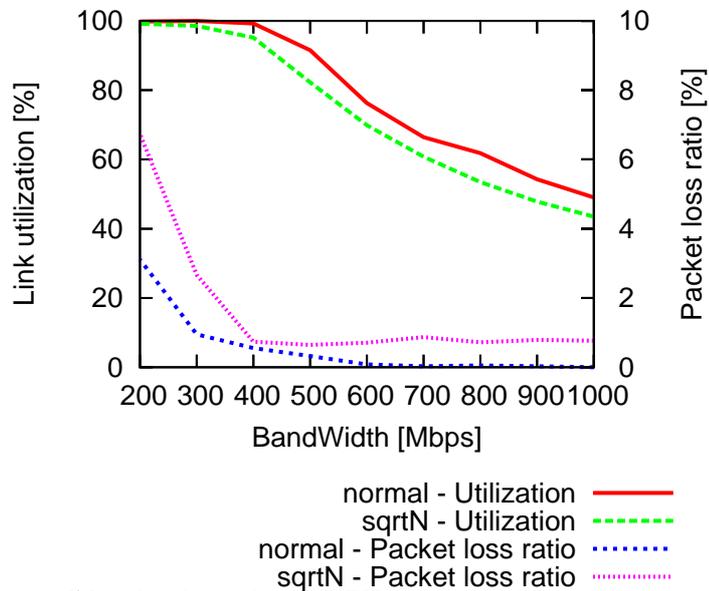
(b) ボトルネックリンクの伝播遅延時間が大きい場合 ($D=90$ ms)

図 7: アクセスリンク帯域がデータ転送遅延時間に与える影響

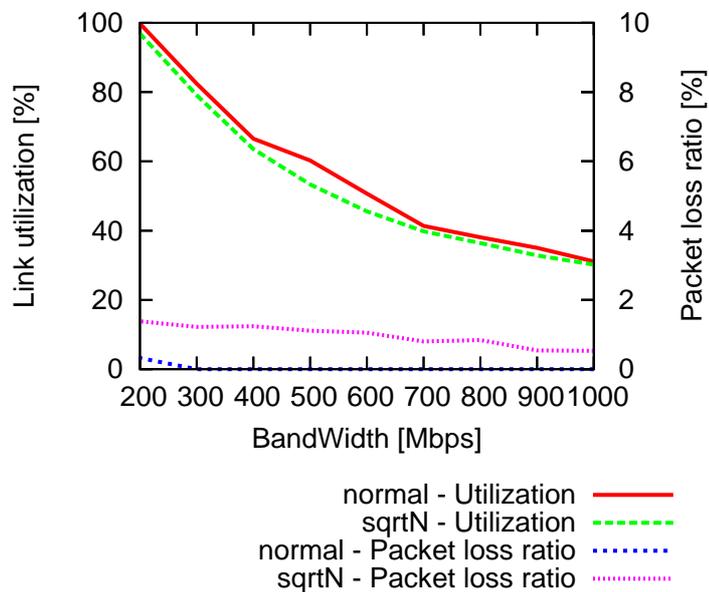
4.2.3 ボトルネックリンク帯域

図 8 は、それぞれ $D=20$ ms、 $N=1500$ および $D=90$ ms、 $N=1500$ の場合において、ボトルネックリンク帯域を変化させた場合のリンク利用率とパケット廃棄率の変化を示している。また、図 9 は、伝送データサイズに対するデータ転送遅延時間の変化を示したものである。図 8 から、リンク利用率、パケット廃棄率のいずれにおいても sqrtN 指標の性能が normal 指標を下回っていることがわかる。特に、ボトルネックリンク帯域が大きい場合においても、パケット廃棄率がある値より小さくならないことが図 8(a) よりわかる。図 8(b) においても、normal 指標ではボトルネックリンク帯域が約 300 Mbps の場合にはパケット廃棄率が 0% になっているが、sqrtN 指標ではリンク帯域が 1 Gbps の場合においてもパケット廃棄が発生している。これは、sqrtN 指標はバッファ容量が小さいため、パケットがバースト的に到着した場合にそれらのパケットをバッファで吸収し切れず、パケット廃棄が発生するためである。また、パケット廃棄率が大きい場合、ボトルネックリンクの伝播遅延時間や転送データサイズが小さい場合においても sqrtN 指標を用いるほうが転送遅延時間が大きくなっている（図 9）。

以上より、リンク帯域が大きく、平均のリンク利用率が低い場合には、sqrtN 指標はリンク利用率、パケット廃棄率、およびデータ転送時間の全てにおいて normal 指標に劣ることが明らかとなった。このことから、始めからネットワーク負荷が低くなるよう設計されている現在のバックボーンネットワークや、ボトルネック帯域が増加しボトルネックリンク利用率が 100% に満たなくなったような場合には、sqrtN 指標を利用することはネットワークを通過する TCP フローに対して性能の低下をもたらすと考えられる。

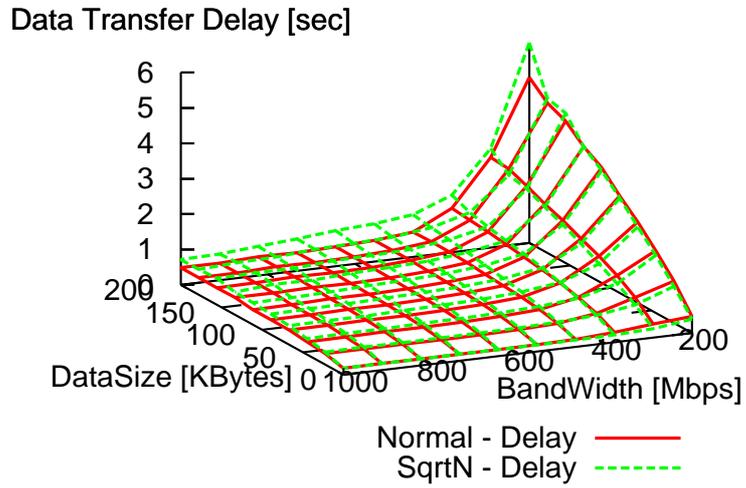


(a) ボトルネックリンクの伝播遅延時間が小さい場合 ($D=20$ ms)

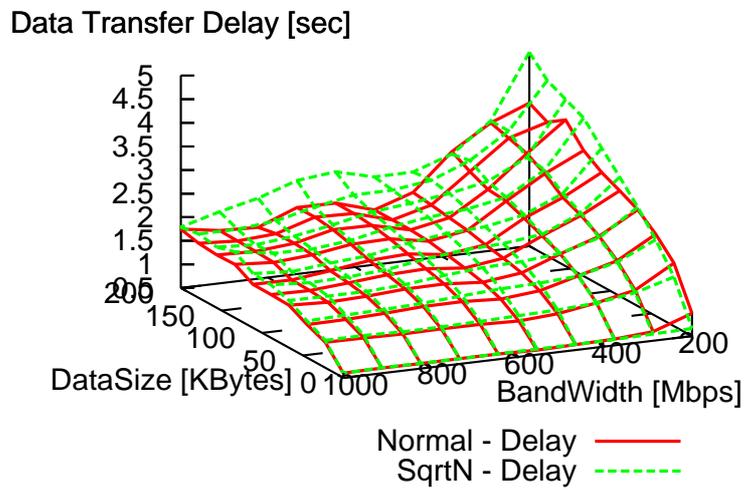


(b) ボトルネックリンクの伝播遅延時間が大きい場合 ($D=90$ ms)

図 8: ボトルネックリンク帯域がネットワーク性能に与える影響



(a) ボトルネックリンクの伝播遅延時間が小さい場合 ($D=20$ ms)

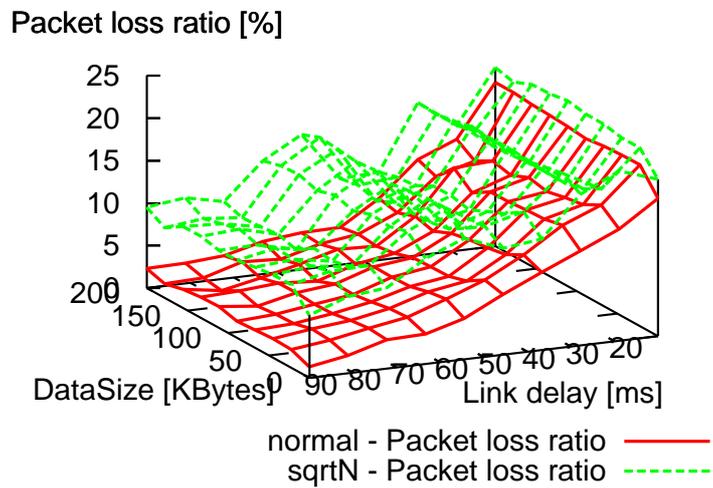


(b) ボトルネックリンクの伝播遅延時間が大きい場合 ($D=90$ ms)

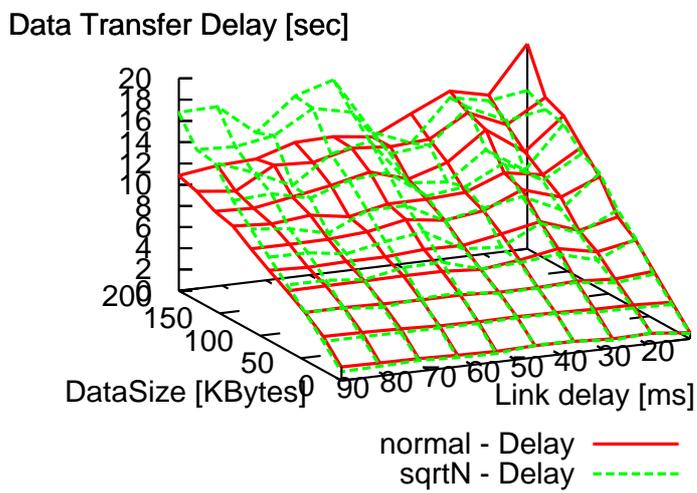
図 9: ボトルネックリンク帯域がデータ転送遅延時間に与える影響

4.2.4 ボトルネックリンクの伝播遅延時間

図 10 は $C = 100$ Mbps、 $N = 1000$ の場合においてボトルネックリンクの伝播遅延時間を変化させたときの、伝送データサイズに対するパケット廃棄率およびデータ転送遅延時間の変化を示したものである。図 10(a) から sqrtN 指標が常にパケット廃棄率が高いこと、およびボトルネックリンクの伝播遅延時間が大きくなるにつれ両指標間のパケット廃棄率の差が大きくなることわかる。図 10(b) においても、ボトルネックリンクの伝播遅延時間が大きくなるにしたがって sqrtN 指標のデータ転送遅延時間が大きくなっている。しかし、sqrtN 指標を利用した場合はパケット廃棄率が高いにもかかわらず、転送データサイズ、あるいはボトルネックリンクの伝播遅延時間が小さい場合には、normal 指標を利用した場合に比べてデータ転送遅延時間が小さい。これは、sqrtN 指標を利用したことでバッファ容量が小さくなった結果、キューイング遅延が減少 (normal 指標に比べて、ホスト数が 900 の場合は 30 分の 1、1600 の場合は 40 分の 1) することにより、パケット廃棄率が高いものの、廃棄されたパケットの再送にかかる時間が短縮されるためであると考えられる。図 10(a) が示す通り、ボトルネックリンクの伝播遅延時間が小さい場合は両指標間でのパケット廃棄率の差が小さくなるため、廃棄されたパケットの再送や輻輳ウィンドウサイズが小さくなることによる影響が縮小する。また、転送データサイズが小さい場合には輻輳ウィンドウサイズが大きくならないため、その減少がデータ転送遅延時間に与える影響は小さい。すなわち、ボトルネックリンクの伝播遅延時間や転送データサイズが小さい場合には、パケット廃棄率よりもキューイング遅延がデータ転送遅延時間に与える影響が大きいため、sqrtN 指標を用いることが、データ転送遅延時間により影響を与えるといえる。



(a) パケット廃棄率



(b) データ転送遅延時間

図 10: ボトルネックリンクの伝播遅延時間が与える影響

5 おわりに

本報告ではシミュレーションを用いて、ルータのバッファ容量が TCP の性能に与える影響の評価を行なった。その結果、sqrtN 指標を用いてバッファ容量を決定した場合は、フロー数が多く平均のリンク利用率が 100% に近い場合は、normal 指標とほぼ同等のリンク利用率が得られる半面、パケット廃棄率が高くなること、ネットワーク負荷が高い場合にはトラヒック量の増加がもたらす影響は小さいが、ネットワーク負荷が低く保たれる場合においては、ボトルネックリンク帯域が大きい場合においてもパケット廃棄率が低下しないため、normal 指標に比べてリンク利用率やデータ転送性能が劣化することが明らかとなった。また、ボトルネックリンクの伝播遅延時間が小さい場合、または転送データサイズが小さい場合を除いて、normal 指標に比べてデータ転送遅延時間に悪影響を及ぼすことがわかった。今後は、ペーシング TCP と TCP-Reno が混在する環境における、バッファ容量が各 TCP の性能に及ぼす影響を考察したい。また、必要なバッファ容量を決定するための重要な要素となる、TCP コネクションが同期して動作するための条件を明らかにしたい。

謝辞

謝辞本報告を終えるにあたり、御教授、御指導を頂きました中野博隆教授に心から感謝いたします。また、同じく御指導頂きました村田正幸教授に心から感謝いたします。並びに本報告の作成を通じて終始熱心に指導、助言を頂きました長谷川剛助教授にも深く感謝いたします。また、的確な助言を頂きました大阪大学情報科学研究科笹部昌弘助手、大阪大学大学院経済学研究科荒川伸一助手に深く感謝いたします。最後に、日ごろから相談に答えて頂きました中野研究室、村田研究室の皆様にご心より御礼申し上げます。

参考文献

- [1] P. Barford and M. Crovella, “Generating representative Web workloads for network and server performance evaluation,” in *Proceedings of the ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pp. 151–160, July 1998.
- [2] J. Postel, “Transmission Control Protocol,” *RFC 793*, Sep. 1981.
- [3] M. Fomenkov, K. Keys, D. Moore, and K. claffy, “Longitudinal study of Internet traffic from 1998-2003,” in *Proceedings of ProWinter International Symposium on Information and Communication Technologies (WISICT 2004)*, Jan. 2004.
- [4] J. Postel, “User Datagram Protocol,” *RFC 768*, Oct. 1980.
- [5] N. Cardwell, S. Savage, and T. Andreson, “Modeling TCP latency,” in *Proceedings of IEEE INFOCOM 2000*, pp. 1742–1751, Mar. 2000.
- [6] W. R. Stevens, *TCP/IP Illustrated, Vol 1: The Protocols*. Reading, Massachusetts: Addison-Wesley, 1994.
- [7] C. Villamizar and C. Song, “High performance TCP in ANSNET,” *SIGCOMM Computer Communications Review*, vol. 24, pp. 45–60, Oct. 1994.
- [8] V. Jacobson, “Modified TCP congestion control algorithm,” *End2end-interest mailing list*, Apr. 1990.
- [9] R. Bush and D. Meyer, “Some Internet architectural guidelines and philosophy,” *RFC 3439*, Dec. 2003.
- [10] G. Appenzeller, I. Keslassy, and N. McKeown, “Sizing router buffers,” in *Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications*, Sep. 2004.
- [11] C. J. Fraleigh, *Provisioning Internet backbone Networks to support latency sensitive applications*. PhD thesis, Stanford University, Department of Electrical Engineering, June 2002. Adviser-Fouad A. Tobagi.
- [12] S. Iyer, R. R. Kompella, and N. McKeown, “Analysis of a memory architecture for fast packet buffers,” in *Proceedings of IEEE High Performance Switching and Routing*, May 2001.

- [13] M. Enachescu, Y. Ganjali, A. Goel, N. McKeown, and T. Roughgarden, "Part III: Routers with very small buffers," *SIGCOMM Computer Communications Review*, vol. 35, pp. 83–90, July 2005.
- [14] S. Floyd and T. Henderson, "The NewReno modification to TCP's cfast recovery algorithm," *RFC 2582*, Apr. 1999.
- [15] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer Networks," *Journal of Computer Networks and ISDN*, vol. 17, pp. 1–14, June 1989.
- [16] S. Shenker, L. Zhang, and D. Clark, "Some observation on the dynamics of a congestion control algorithm," *ACM Computer Communications Review*, vol. 20, pp. 30–39, Oct. 1990.
- [17] Y. J. A. Gilbert and N. McKeown, "Congestion control and periodic behavior," in *Proceeding of LANMAN Workshop*, Mar. 2001.
- [18] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397–413, 1993.
- [19] L. Zhang and D. Clark, "Oscillating behavior of network traffic: A case study simulation," *Internetworking: Research and experience*, vol. 1, no. 2, pp. 101–112, 1990.
- [20] L. Qiu, Y. Zhang, and S. Keshav, "Understanding the performance of many TCP flows," *Computer Networks*, vol. 37, no. 3-4, pp. 277–306, 2001.
- [21] G. Iannaccone, M. May, and C. Diot, "Aggregate traffic performance with active queue management and drop from tail," *ACM Computer Communication Review*, vol. 31, pp. 4–13, July 2001.
- [22] T. V. Project, "UCB/LBNL/VINT network simulator - ns (version 2)." available at <http://www.isi.edu/nsnam/ns/>.