

VoIP の性能評価と待ち行列モデル

滝根 哲哉

京都大学大学院情報学研究科数理工学専攻

1 はじめに

インターネットの高速化と利用者の増大に伴い、インターネットにおける実時間制約をもつアプリケーションが注目されている。実時間制約をもつアプリケーションを快適に動作させるには通信サービス品質 (QoS: Quality of Services) の確保が重要となる。本稿では特にインターネットにおける音声通信に焦点をあて通信サービス品質を評価するための道具としての待ち行列モデルを紹介する。

音声トラフィックは異なる回線速度をもつドメイン網、アクセス網、スタブ網の三つのタイプのネットワークを経由して伝送される。インターネットのバックボーンを構成する ISP (Internet Service Provider) が管理するドメイン網は高速の回線を持っているが、ドメイン網へアクセスするための回線群であるアクセス網は比較的低速なものとなっている。また、スタブ網は大学や大きな建物に施設された LAN (Local Area Network) に対応しており、小規模ながら高速な回線を持つ。

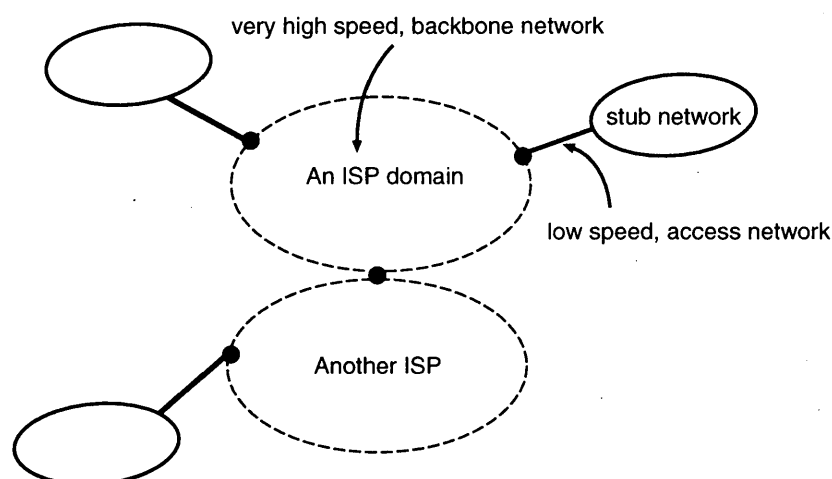


図 1: ネットワーク構成

インターネット上で QoS を確保するための基本的な機構については IETF (Internet Engineering Task Force) によって、Integrated Services (IntServ) や Resource Reservation Protocol (RSVP) といったものが標準化されているが、まだ未解決の問題が二つ残されている。一つは、低速なアクセス網において適当な QoS を提供するために遅延に課せられた制約を如何に満たすかということである。アクセス網で同時に扱う音声フロー数は比較的少ないため、多少、手の込んだ制御であっても実装可能と考えられる。もう一つの問題はどのようにして高速なドメイン網を流れる多数の音声フローを扱うかという問題である。例えば 10Gb/s の回線は 10kb/s の音声フローを最大 1,000,000 まで収容できる。高速網では、適当な帯域割り当て制御を行うことで網内の遅延に関しては無視できる考えられるため、そのような帯域割り当てを大量の音声フローに対して如何に行うかが主な問題となる。

以下では上記二つの問題を考える。まず初めにアクセス網で音声フローが被る遅延を評価するためのモデルを考察する。アクセス網では音声フローのみならずデータなどのトラフィックが混在していると考えら

れる。そこで、音声フローが要求する QoS を満たすため、音声フローが他のトラフィックに対して非割り込み優先権を持たせることを想定する。このようなシナリオの下で音声フローの遅延分布関数の導出ならびに非優先フローの packets 長が遅延分布に与える影響について考察する。続いてドメイン網での帯域割り当て制御法として複数の音声フローを収容するトランクという概念を導入し、トランク単位で帯域を割り当てる手法の性能を評価するためのモデルを考察する。

2 アクセス網における遅延性能評価モデル

前章で述べたようにアクセス網では将来に渡って比較的低速な回線が利用されることが予想されるため、ここでの遅延が音声フローの QoS を決める重要な要素となる。現在、この問題は IETF IntServ-specific link layers (ISSLL) working group で議論が行われており、音声パケットの遅延を減少させる手法として、ヘッダ圧縮とセグメント化の二つの提案がなされている。ヘッダ圧縮は音声パケットに適用され、低速な回線における送信時間の短縮を目的としている。一方、セグメント化は非音声パケットに適用され、パケットを小さい単位に分割することで、優先権をもつ音声パケットへの影響を小さくする目的で用いられる。以下では、これらの方策がどの程度有効かを定量的に評価するための待ち行列モデルを考える。

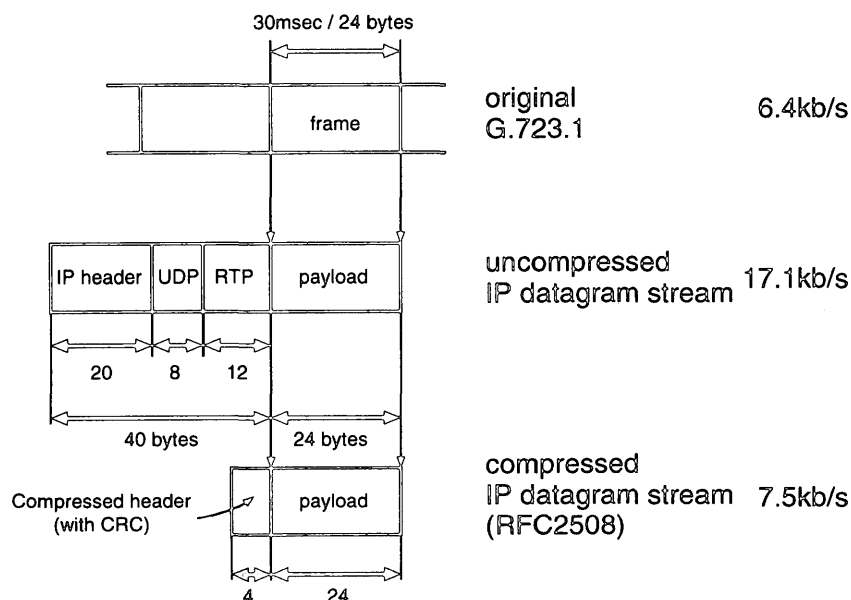


図 2: 音声パケットヘッダの圧縮

N 本の音声フローと非音声フローがアクセス回線を共有していると仮定する。それぞれの音声フローは、時間間隔 D 毎に周期的に伝送時間 b の固定長パケットを送出する。ただしそれぞれのフローの初期位相は互いに独立に $(0, D]$ の一様分布に従うと仮定し、 $Nb < D$ であるとする。音声パケットは非音声パケットに対して非割り込み優先権をもっていると仮定する。音声パケット間でのサービス順序は FIFO とする。また、非音声パケットの伝送時間は一定で u であるとし、伝送すべき音声パケットがない場合は直ちに非音声パケットの伝送が開始されるとする。¹

このモデルは 休暇付き $nD/D/1$ 待ち行列である。詳細は省略するが、 W を音声パケットの待ち時間

¹最後の仮定は、音声パケットから見た時の最悪な場合に対応している。

を表す確率変数とし、 $\bar{W}(x) = \Pr(W > x)$ を待ち時間の補分布とすると、次式を得る。

$$\begin{aligned} \bar{W}(x) = \frac{D}{Nu} & \left[\sum_{n=0}^{N-1} \left(\frac{\alpha_{nb+u}(x)}{D} \right)^{n+1} \sum_{k=0}^{N-n-1} \binom{k+n}{n} \left(1 - \frac{\alpha_{nb+u}(x)}{D} \right)^k \right. \\ & \left. - \sum_{n=0}^{N-1} \left(\frac{\alpha_{nb}(x)}{D} \right)^{n+1} \sum_{k=0}^{N-n-1} \binom{k+n}{n} \left(1 - \frac{\alpha_{nb}(x)}{D} \right)^k \right] \\ & - \frac{b}{u} \left[\sum_{n=0}^{N-2} \left(\frac{\alpha_{nb+u}(x)}{D} \right)^{n+1} \sum_{k=0}^{N-n-2} \binom{k+n}{n} \left(1 - \frac{\alpha_{nb+u}(x)}{D} \right)^k \right. \\ & \left. - \sum_{n=0}^{N-2} \left(\frac{\alpha_{nb}(x)}{D} \right)^{n+1} \sum_{k=0}^{N-n-2} \binom{k+n}{n} \left(1 - \frac{\alpha_{nb}(x)}{D} \right)^k \right]. \end{aligned}$$

この式は $\max(0, (N-1)b+u-D) \leq x \leq (N-1)b+u$ なる x に対して数値的に安定である。

図 3 は、回線速度 1.536 Mb/s、音声パケット長 28 byte（内ヘッダ 4 byte）、符号化速度 6.4kb/s、非音声パケット長 500byte の場合の負荷 Nb/D に対する待ち時間分布の 99 パーセントイル、最悪値ならびに平均待ち時間を示したものである。これより、99 パーセントイルは高々平均の数倍程度であるのに対し、最悪値は負荷に線形に増加することが分かる。よって確率的受け付け制御を行うことでより多くの音声パケットが収容可能であることが分かる。

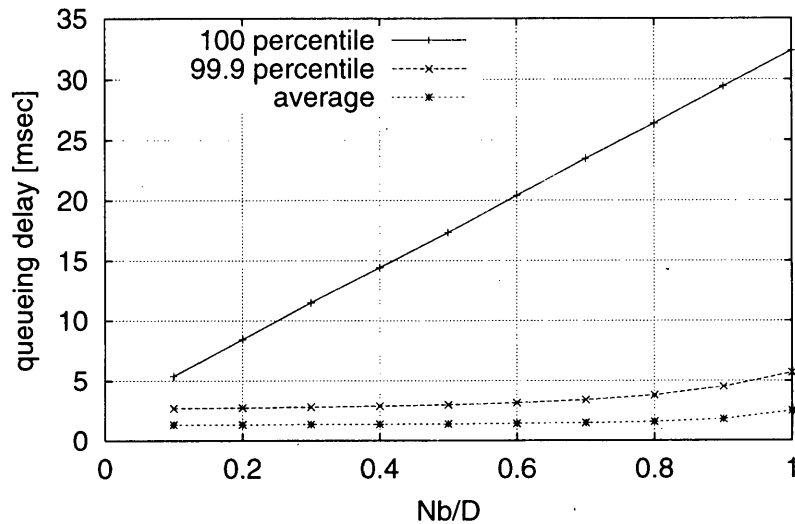


図 3: 待ち時間分布

図 4 は、回線速度 128kb/s、 $N = 7$ 、符号化速度 6.4kb/s の場合にヘッダ圧縮を行わない場合（音声パケット長 68 byte、内ヘッダ 40 byte）とヘッダ圧縮を行った場合（音声パケット長 28 byte、内ヘッダ 4 byte）の待ち時間分布の 99 パーセントイルを非音声パケットのサイズの関数として表したものである。この図より、ヘッダ圧縮によって 99 パーセントイル待ち時間が 10ms 程度減少すること、ならびに 99 パーセントイル待ち時間が非音声パケット長に対してほぼ線形の関数となることが分かる。

3 バックボーン網におけるフロー管理

インターネットにおいて実時間制約をもつアプリケーションを収容する方法に関しては IntServ/RSVP が標準化されている。しかし RSVP はフロー毎の管理を前提としているため、大規模化に対応できないこ

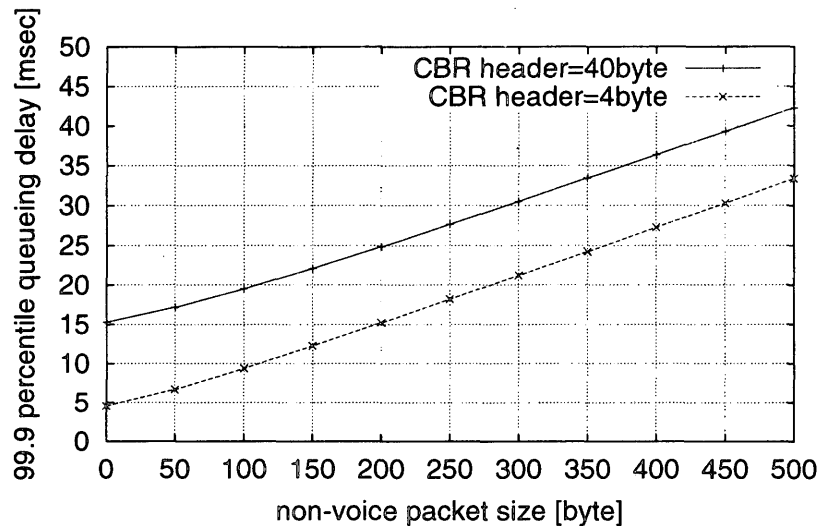


図 4: 99 パーセントイル待ち時間

とが指摘されている。この問題を解決するため IETF では Differentiated Services (DiffServ) の標準化が進められている。DiffServ では「網端」²という概念が導入されており、網端においてパケットに対するマーク付けならびにトラフィック整形が行われる。一方、高速なバックボーン網を構成している網内のルータでは各パケットをマークによってのみ差別化を行う。この差別化は PHB (Per-Hop Behavior) と呼ばれている。この結果、DiffServ ではフロー集約が行われることになる。すなわち、ドメイン網の内部では同じマークをもつ多数のフローはまとめられ、単一のフローとして扱われることになる。

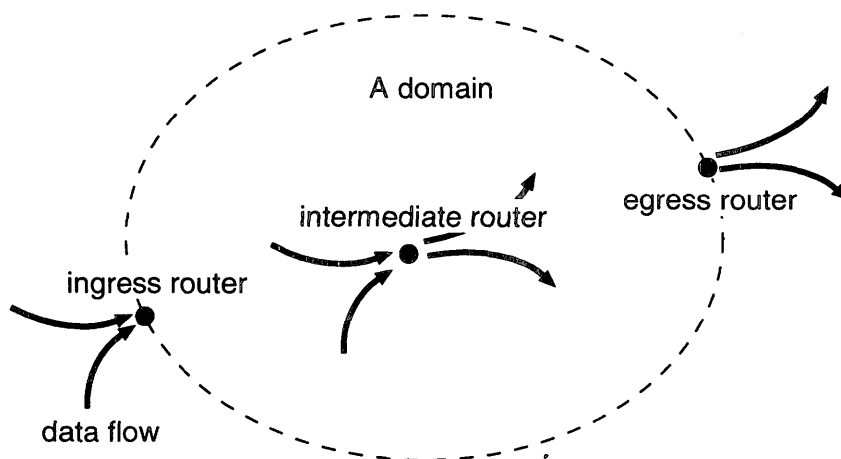


図 5: フロー集約とルータの役割

以下では音声フローが集約された場合の性能とフロー管理のためのオーバーヘッドについて入口ルータに注目し考察する (図 6 参照)。入口ルータにおける呼の振舞いをモデル化すると図 7 に示すような、2 ステージ呼損モデルが得られる。ここでトランクとは多くの音声フローが集約されている 1 つの管理単位であり、網内のルータはこのトランクのみを認識している。ある音声フローの設定要求が入口ルータで発生したとき、このフローを収容するトランクが利用可能でなく、帯域に余裕がある場合は、出口ルータまでの

²通常、ISP が管理するドメインの端点を指す。

経路に沿ってトランクが確保される。もし、帯域に余裕がない場合はこの設定要求は却下される。さらに、トランクが確保されると、トランク内では最大 m 本の音声フローを収容することが出来る。トランク内に既に m 本のフローが収容されている時に発生した新たな音声フローの設定要求は棄却される。この結果、呼損の発生はフローが確立されているが新たなフローを収容できない場合と、新たなトランクの確保が出来ない場合の二通りの場合が起こる。

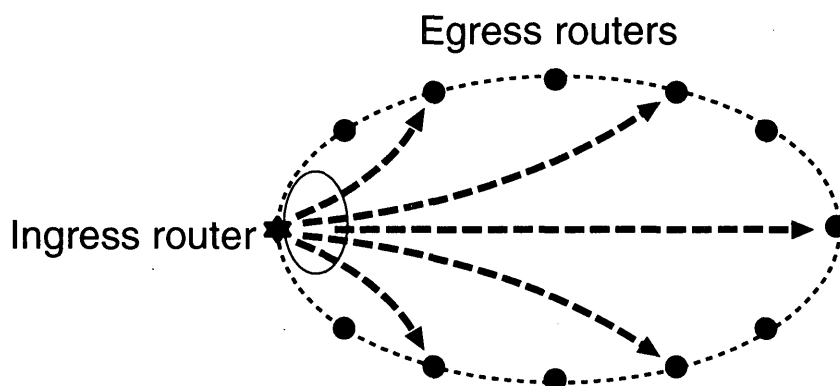
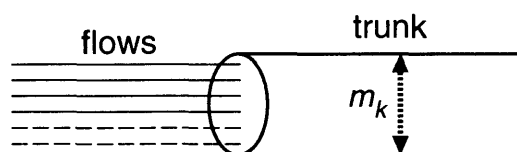


図 6: 対象システム

The first stage:
blocking is caused by contentions among flows in a trunk



The second stage:
blocking is caused by contentions among trunks

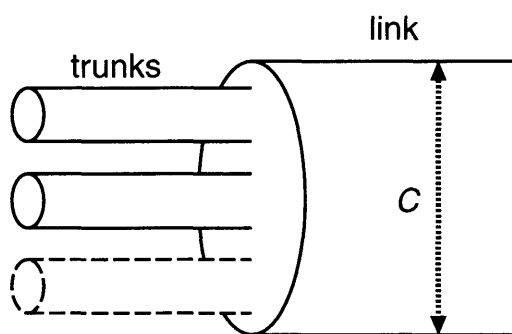


図 7: 2 ステージ呼損モデル

音声フローをトランク単位で扱うため、フロー単位の管理と比較して、管理しなければならない情報量はおよそ $1/m$ になる。しかし、トランク内の帯域が完全に利用されているわけではないため無駄が生じ、結果として呼損率は悪化するものと考えられる。以下では、各トランクへの音声フローの発生はポワソン過程に従い、各音声フローの保留時間は独立同一分布に従うと仮定して、以下のような 2 ステージモデルの呼損率を求める。

トランクは収容する呼の性質によって P 個のタイプに分類される。タイプ k のトランクは N_k 本あり、これらのトランクのそれぞれへの呼の発生は率 λ_k のポワソン過程に従う。タイプ k のトランクは m_k 個の呼を収容できるとし、2 段目における回線は合計 C 個の呼が収容できると仮定する。

$\rho_k = \lambda_k h_k$ を一つのタイプ k のトランクへの負荷とする。このとき、一つのタイプ k のトランク内に少なくとも 1 個の呼が収容されているという条件の下でのタイプ k のトランク内に収容されている呼の平均数 M_k は次式で与えられる。

$$M_k = \frac{\rho_k(1 - B(m_k, \rho_k))}{1 - P_k(0)}$$

ただし、 $P_k(0)$ は $C = \infty$ の場合におけるタイプ k のトランクが空である確率であり

$$P_k(0) = \left[\sum_{i=0}^{m_k} \frac{\rho_k^i}{i!} \right]^{-1}$$

で与えられる。また $B(m, \rho)$ は負荷が ρ である $M/G/m/m$ における呼損率であり、 $B(0, \rho) = 1$ ならびに

$$B(m, \rho) = \frac{\rho B(m-1, \rho)}{m + \rho B(m-1, \rho)}, \quad m = 1, 2, \dots$$

によって計算できる。また、タイプ k の空のトランク内に 1 個の呼が収容されてから再び空になるまでの平均時間 H_k は

$$H_k = \frac{1 - P_k(0)}{\lambda_k P_k(0)}$$

で与えられる。

以上の準備の下で呼損率を考える。 $Q(\mathbf{n})$ $\mathbf{n} = (n_1, \dots, n_P)$ を n_k ($k = 1, \dots, P$) 本のタイプ k のトランクが回線を確保している定常確率とする。このとき

$$Q(\mathbf{n}) = \begin{cases} \frac{\prod_{k=1}^P \binom{N_k}{n_k} \sigma_k^{n_k}}{\sum_{\mathbf{n} \in \mathcal{T}(\mathbf{N}, C)} \prod_{k=1}^P \binom{N_k}{n_k} \sigma_k^{n_k}}, & \mathbf{n} \in \mathcal{T}(\mathbf{N}, C) \\ 0, & \text{otherwise} \end{cases}$$

となる。ただし $\mathbf{N} = (N_1, \dots, N_P)$ であり、 σ_k ならびに $\mathcal{T}(\mathbf{N}, C)$ は

$$\sigma_k = \lambda_k H_k, \quad k = 1, \dots, P$$

$$\mathcal{T}(\mathbf{N}, C) = \left\{ \mathbf{n} \mid \sum_{k=1}^P n_k m_k \leq C, 0 \leq n_k \leq N_k \ (k = 1, \dots, P) \right\}$$

で与えられる。

$U_k(\mathbf{N}, C)$ を回線を確保しているタイプ k の平均トランク数とする。

$$U_k(\mathbf{N}, C) = \sum_{\mathbf{n} \in \mathcal{T}(\mathbf{N}, C)} n_k Q(\mathbf{n})$$

このときタイプ k のトランクに収容される呼の呼損率 $P_{loss}(k)$ は

$$P_{loss}(k) = 1 - \frac{M_k U_k(\mathbf{N}, C)}{N_k \rho_k}$$

で与えられる。また、1 段目で呼損となる確率を $P_{loss}^{(1)}(k)$ とすると

$$P_{loss}^{(1)}(k) = \frac{U_k(\mathbf{N}, C)}{N_k} \cdot \frac{B(m_k, \rho_k)}{1 - P_k(0)}$$

で与えられる。よって 2 番目で呼損となる確率 $P_{loss}^{(2)}(k)$ は $P_{loss}(k) - P_{loss}^{(1)}(k)$ で与えられる。

図 8 はトランクのタイプが一種類の場合の呼損率を示したものである。

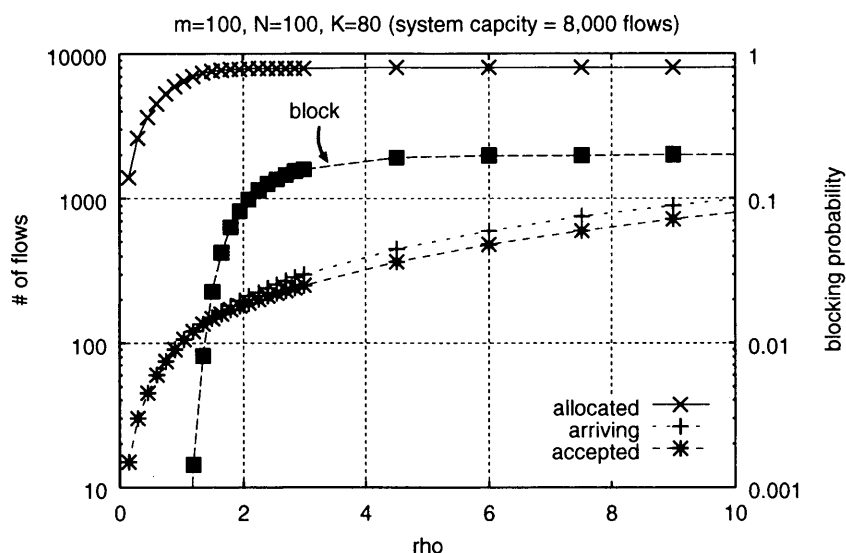


図 8: 呼損率

最初に述べたように m の値は管理のオーバーヘッドと密接な関係があり、 m を増加させることでオーバーヘッドを $1/m$ にすることが出来る。しかし、その対価として性能が劣化する。しかしトランクの容量が大きくなると大群化効果によってその差は小さくなる。表 1 は各フローの速度が 10Kb/s で $N = 100$ の場合に、呼損率を 10^{-4} に押えるために必要なリンク容量を、トランクを用いないシステムで同じ呼損率を達成するために必要な容量と比較したものである。表より、大規模なシステムではトランクを導入することで必要となる余分な帯域は相対的に小さくなる事が分かる。

表 1: 大群化効果

| ρ | Mean # of flows can be accepted | Ratio of the provisioned capacity of aware system to that of unaware system |
|-----------|---------------------------------|---|
| 10Mb/s | 1,000 | $\frac{24.0\text{Mb/s}}{11.0\text{Mb/s}} = 2.18$ |
| 100Mb/s | 10,000 | $\frac{137\text{Mb/s}}{103\text{Mb/s}} = 1.33$ |
| 1,000Mb/s | 100,000 | $\frac{1.10\text{Gb/s}}{1.01\text{Gb/s}} = 1.09$ |

参考文献

- [1] K. Iida, T. Takine, H. Sunahara and Y. Oie, "Delay Analysis for CBR Traffic in Class Based Queueing: Single-Node and Homogeneous CBR Traffic Case," *Proceedings of SPIE Conference of Performance and Control of Networks*, vol.3231, pp.311-322, Dallas, TX, November 3-7, 1997.
- [2] K. Iida, T. Takine, H. Sunahara and Y. Oie, "Delay Analysis for CBR Traffic in Static-Priority Scheduling: Single-Node and Heterogeneous CBR Traffic Case," *Proceedings of IEEE GLOBECOM '98*, pp.1256-1263, Sydney, Australia, November 8-12, 1998.
- [3] K. Iida, K. Kawahara, T. Takine and Y. Oie, "Performance Analysis of Flow Aggregation of Constant Bit Rate Type Traffic at Ingress Router" *Proceedings of IEEE GLOBECOM '99*, pp.92-99, Rio de Janeiro, Brazil, 5-9 December, 1999.

- [4] K. Iida, K. Kawahara, T. Takine and Y. Oie, "Performance Evaluation of the Architecture for End-to-End Quality-of-Service Provisioning," *IEEE Communications Magazine*, vol.38, no.4, pp.76–81, 2000.
- [5] K. Iida, T. Takine, H. Sunahara and Y. Oie, "Delay Analysis for CBR Traffic in Multimedia Enterprise Network," to appear in *IEICE Transactions on Communications*, 2001.
- [6] K. Iida, T. Takine, H. Sunahara and Y. Oie, "Delay Analysis for CBR Traffic under Statistic-Priority Scheduling," to appear in *IEEE/ACM Transactions on Networking*, 2001.