Adaptive Algorithms for Feedback-Based Flow Control in High Speed, Wide Area ATM Networks

Flavio Bonomi[*] *Debasis Mitra*[†] *Judith B. Seery*[†]

[*]ATM Platform Org., AT&T Network Systems, Red Bank, NJ 07701

[†]AT&T Bell Laboratories, Murray Hill, NJ 07974

Abstract

This paper gives a class of flow control algorithms for the adaptive allocation of bandwidths to virtual connections (VC) in high speed, wide area ATM networks. The feedback rate to the source from the network is parsimonious, with each feedback bit indicating whether the buffer at a distant switch is above or below a threshold. The service discipline at the switch is First-Come-First-Served. The important goal of adaptability aims to make all of the network bandwidth available to the active VCs, even though the number of such VCs is variable over a given range. Each VC has two parameters, one giving its minimum guaranteed bandwidth and the other is the weight for determining its share of the uncommitted bandwidth. Judicious selection of these parameters defines distinctive services, such as Best Effort and Best Effort with Minimum Bandwidth. We derive design rules for selecting the parameters of the algorithms such that the appropriate guarantees and fairness properties are exhibited in the dynamical behavior. The systematic use of "damping" in right proportion with "gain" is shown to be a powerful device for stabilizing behavior and achieving fairness. Our analyses are based on a simple analytic fluid model composed of a system of first-order delay-differential equations, which reflect the propagation delay across the network. Extensive simulations examine the following: (i) fairness, especially to start-up VCs; (ii) oscillations; (iii) transient behavior, such as the rate of equalization from different initial conditions; (iv) disparate bandwidth allocations; (v) multiple paths with diverse propagation delays; (vi) adaptability and robustness with respect to parameters; (vii) interoperability of different algorithms.

1 INTRODUCTION

Feedback-based flow control in wide-area networking suffers from the unavoidable fact that the time constants associated with the adaptive processes are of the order of the large propagation delays, which, when coupled with high speeds, translates to a major limitation. On the other hand, there is a certain broad class of bursty application processes such that, when the demand for bandwidth arises, it persists for time periods which are comparable to the aforementioned time constants or longer. For users of such processes, feedback-based flow control offers value. Also, besides the direct goal of allocating bandwidth, feedback algorithms are in the related business of providing "backward information propagation", i.e., communicating allocations from the network to the user. Without such information, the user cannot fully exploit the bandwidth allocated to it, no matter how generous, and how efficient and fair the process. These and other reasons help to explain the wide interest and the intensity of recent proceedings of the ATM Forum dealing with the Available Bit Rate service category and the role of feedback to support it. The interested reader may consult [BFE95] and references therein.

This paper gives a class of flow control algorithms, which, while quite general by design, is primarily intended for the environment where feedback to the user from the network is parsimonious and in the form of single bits. In the base case the rate of the feedback stream is approximately uniform with the interarrival time between bits a parameter, which may range typically from about 0.1 to 1.0 times the round-trip propagation delay, τ. We envisage the algorithms being applied to virtual connections (VC) in wide area networks, perhaps spanning the continental USA, and hence propagation delays may be large. The feedback information conveyed by each bit is simply whether the instantaneous queue length at the distant switch is above a threshold or not. Thus the feedback is explicit [JAI90]. See [RJA88] and [JAC88] for important early work on adaptive, feedback-based algorithms.

Figure 1.1 is a sketch of the network model considered here. To keep matters simple, we consider only a single switch in the paths of the various VCs. The figure shows various propagation delays for the VCs; hence the sources may be located at various geographical sites. The figure shows μ and Q_T to be, respectively, the service rate and the queue threshold at the distant node. The packet or cell size is assumed to be fixed, not randomly distributed. Variability in this study is due solely to the dynamics of the adaptation and the turning on and off of the VCs.

Our main contributions are as follows. We formalize certain desirable properties of bandwidth allocations which deal with fairness and guarantees on minimum bandwidth, and identify dynamical behavior which converges to give such properties. We give a unified framework for designing a broad class of dynamical feedback algorithms which satisfy the aforementioned properties. We show that the systematic use of "damping" is a powerful device for stabilizing behavior and for achieving fairness; when damping is used in the right proportion with "gain", then the disadvantages of damping, namely, reduced dynamic range and sluggish transient behavior, may be effectively nullified. Finally, we demonstrate

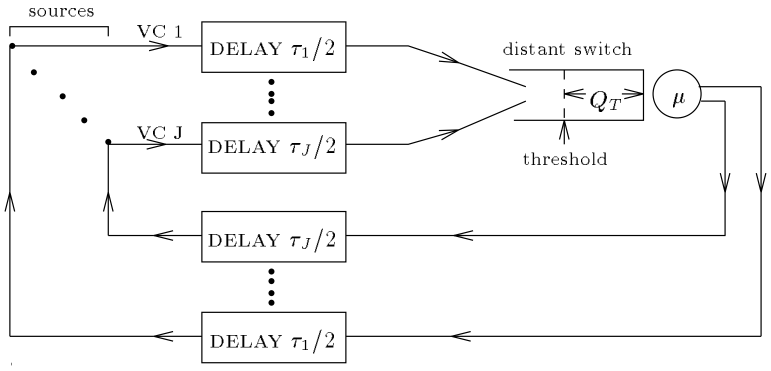

Figure 1.1: Network model.

through simulations that, when properly designed, the algorithms demonstrate adaptability and performance, which is fair, stable and responsive.

Each VC has two nonnegative parameters associated with its bandwidth allocation, ν_j and σ_j, for virtual connection j. The rate parameter ν_j is the minimum bandwidth and σ_j is the positive weight given to the VC in determining its share of the uncommitted bandwidth. Various services with distinctive characteristics may be designed by appropriate choice of these parameters. Best Effort and Best Effort with Minimum Bandwidth are examples of such services given in Section 4.

The algorithms given in this paper are for VCs which share a common buffer at the network node and the service discipline at the node is First-Come-First-Served. (In separate work we have examined various weighted round robin disciplines, where fairness is more easily obtained.) An important reason for considering the shared memory – FCFS implementation is its simplicity and low cost. Similarly, in today's technology, processing capacity is at a premium, and parsimonious feedback together with a relatively low rate of updates, is highly desirable for its low processing requirements.

The analyses in Sections 3 and 4 are based on an analytic fluid model which approximates the algorithms. The analytic results provide guidelines for the design of parameters of the algorithms. The results show that there are two distinct operating regimes, Unsaturated and Saturated. In steady state, the former has unutilized network bandwidth, oscillation-free flow and empty buffers. In the latter regime all bandwidth is utilized, and both flow and queue behavior oscillates around μ and Q_T, respectively. An important attribute of the algorithms is that as the degree of saturatedness increases, as happens with increasing number of VCs, the amplitude and energy of the oscillations increase gracefully. That is, there is a very broad range of parameters in the Saturated regime for which performance is acceptable in terms of high utilization and low delay. This robustness is at the core of the adaptability of the algorithms. That is, when the number of virtual circuits J is allowed to vary over a broad range and parameters of ongoing connections are not rescaled with J, performance goals continue to be satisfied.

An important goal is to make all of the network bandwidth available to the active VCs, with the individual allocations commensurate with their allocation parameters. Our simulation studies which are reported in Section 5 make the assumption that when a VC turns on it acts like an infinite data source, i.e., with unbounded data to send. Hence, even when there is only a small number of VCs on, the utilized fraction of the bandwidth is close to unity, and this is also true at the other extreme when the maximum number of permissible VCs are on simultaneously.

We briefly review some prior work. The algorithms proposed and designed in [MSE90, MSE93(i), MIT92, FRW92] were based on packet-by-packet adaptation and driven by response time measurements. In both these respects the present work is different. Elwalid [ELW93] has developed a fundamental theory, based on delay-differential equations, of response-time based adaptations in the presence of large propagation delays. He shows how damping and gain parameters may be selected to optimize transient behavior. [MSE93(ii)] first gave the basic symmetric algorithm discussed later in this paper, and also the design rules for the algorithm.

We list some key issues which are examined in this paper, as well as in various other prior studies. (i) Fairness in short and long term behavior. This is specially important for a start-up VC, which has difficulty in obtaining its share of the bandwidth from existing VCs. (ii) Large amplitude oscillations. Zhang [ZHA89] has shown that they exist in certain algorithms. Fendick et al. [FRW92] give a beautiful treatment in which they prove that such oscillations are controlled to acceptably small magnitudes in a class of algorithms. Mukherjee and Strikwerda's analysis [MST91] illuminates the possibility of large amplitude oscillations in the JRJ algorithm. (iii) Transient behavior, i.e. the quality of the responsiveness in a dynamic environment. Mishra and Kanakia [MKA92], Keshav [KES91], Bolot and Shankar [BSH90], Mackrucki [MAC91], Benmohamed and Meerkov [BME92], Hluchyj and Yin [HYI94], Kanakia et al. [KMR93] and Keshav [KES94] provide studies in which algorithms are proposed and investigated with the focus on transient behavior. (iv) Arbitrary bandwidth allocation to VCs. (v) Various propagation delays. Zhang [ZHA89] has shown that there is typically a tendency to discriminate against VCs on longer paths. (vi) Robustness with respect to parameters.

2 ALGORITHMS

In this section we begin by giving fairly general asynchronous versions of the algorithms for rate control. Next, the algorithm is specialized to some specific implementations and to the synchronous version, which is followed by some simple enhancements. Finally, the counterpart for windows' control is described.

2.1 Rate Control

In the description below $R_j(t)$ denotes the rate allocation to the j^{th} VC just after an update at time t. The algorithm is asynchronous, and hence the VCs which have their rates updated simultaneously at t form an indeterminate set. Let $U(t)$ denote this set. The number of elements in $U(t)$ may range from 1 to J, where J is the total number of VCs.

$$R_j(t) = \begin{cases} R_j(t-d_j(t)) - \gamma_j(u_j(t))d_j(t)[R_j(t-d_j(t)) - \nu_j] + a_j(u_j(t))d_j(t)u_j(t) & \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad j \in U(t) \\ R_j(t-0) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if} \quad j \notin U(t) \end{cases} \tag{2.1}$$

Here $u_j(t) \in \{-1, 1\}$ is the feedback received at the source at time t just prior to the update in the rate allocation, from the distant switch. An update may be triggered by various events, including the expiration of a timer, a counter reaching a predetermined value, or the receipt of feedback information. As indicated in Figure 1.1 it is assumed that the propagation time between source and distant switch for VC$_j$ is $\tau_j/2$ in each direction. This assumption is particularly justified in wide area networks, where the propagation delay far exceeds other components of the transport delay. Some results on the robustness of our rate control algorithms with respect to this assumption will be presented in Section 5. The feedback bit $u_j(t)$ is $+1$ or -1 depending upon whether the queue at the distant switch at time $(t-\tau_j/2)$, $Q(t-\tau_j/2) < Q_T$, or $\geq Q_T$, respectively. The rate parameter ν_j in (2.1) is the user's expectation of minimum bandwidth. Also, $\gamma_j(u)$ is the damping function and $a_j(u)$ is the gain function, which are nonnegative design parameters of the algorithm and the subject of discussions later in the paper. We only consider the following special forms of these functions:

$$\gamma_j(u) = \begin{cases} \gamma_j^+ & \text{if} \quad u > 0 \\ \gamma_j^- & \text{if} \quad u < 0 \end{cases} \quad \text{and} \quad a_j(u) = \begin{cases} a_j^+ & \text{if} \quad u > 0 \\ a_j^- & \text{if} \quad u < 0 \end{cases} \tag{2.2}$$

Particular attention is given to two algorithms, which are specified by the following choice of parameters.

Jacobson-Ramakrishnan-Jain (JRJ) [JAC88, RJA88]:

$$\begin{aligned} &\gamma_j^+ = a_j^- = 0 \\ &\text{Let } \gamma_j^- = \gamma_j \text{ and } a_j^+ = a_j. \end{aligned} \tag{2.3}$$

Mitra-Seery (MS) [MSE93(ii)]:

$$\begin{aligned} \gamma_j^+ &= \gamma_j^- \triangleq \gamma_j \\ a_j^+ &= a_j^- \triangleq a_j. \end{aligned} \tag{2.4}$$

Thus the JRJ algorithm is asymmetric in the expansionary and contractive phases, in contrast to the symmetry of the MS algorithm, which is uniformly damped.

In (2.1) $d_j(t)$ is the "delay", or time interval, between the prior update of the rate of VCj and the update at time t.

The asynchronous nature of the algorithm in (2.1) allows substantial generality in implementations of feedback control. By specializing (2.1) various specific updating and congestion detection schemes may be accommodated. For instance, ATM Resource Management (RM) cells may be used to carry back congestion information to the source. These RM cells may be transmitted periodically, with the period determined by the geographical length of the virtual connection. Another scheme, which too is described by (2.1), is where a RM cell is transmitted for every N data cells transmitted. In the latter scheme the overhead traffic is kept to a fixed proportion of the useful data traffic, and has the attraction of scalability in the framework of many small virtual connections. The reader may consult [BFE95] for further information on these and other schemes.

The asynchronous model in (2.1) may be specialized to the following synchronous model, which is itself widely applicable and also helpful as an intermediate between (2.1) and the fluid models of succeeding sections. Let Δ_j denote the fixed interval between updates of the rate for VC_j, and abbreviate $R_j(n\Delta_j)$ to $R_j(n)$, $n = 0, 1, \cdots$. Then,

$$R_j(n) = R_j(n-1) - \gamma_j(u_j(n))\Delta_j[R_j(n-1) - \nu_j] + a_j(u_j(n))\Delta_j\, u_j(n), \tag{2.5}$$

where $u_j(n\Delta_j)$ has also been abbreviated to $u_j(n)$.

2.2 Floor and Ceiling

We modify (2.1) to incorporate two new mechanisms, "Floor" and "Ceiling". First, define the function $\mathrm{sat}_j(\cdot)$ thus:

$$\begin{aligned} \mathrm{sat}_j(x) &= x && \text{if} \quad \nu_j \le x \le R_j^{\mathrm{ceil}} \\ &= \nu_j && \text{if} \quad x < \nu_j \\ &= R_j^{\mathrm{ceil}} && \text{if} \quad x > R_j^{\mathrm{ceil}}, \end{aligned} \tag{2.6}$$

where R_j^{ceil} is a parameter of VCj. The modification to (2.1) is,

$$R_j(t) = \begin{cases} \mathrm{sat}_j[R_j(t-d_j(t)) - \gamma_j(u_j(t))d_j(t)\{R_j(t-d_j(t)) - \nu_j\} + a_j(u_j(t))d_j(t)u_j(t)] \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad j \in U(t) \\ R_j(t-0) \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{if} \quad j \notin U(t) \end{cases} \tag{2.7}$$

The motivation for Floor is that it guarantees that at all points in time the rate allocated to VCj is not less than ν_j. It's main effect is during start-up. As will be seen later, even without Floor the algorithms dynamically converge to satisfy this rate requirement.

The motivations for Ceiling are different. One such motivation arises when the design is for possibly many VCs, but only a few VCs are active. Ceiling restricts the rates of these active VCs, so that the surge in the buffer at the distant node is mitigated when several VCs turn active at about the same time.

2.3 Window Control

The window for a VC restricts the number of packets or cells which have been transmitted but not acknowledged as having been received by the distant switch. Windows and rates are closely related. Our simulation studies have covered both. A major difference in our algorithms is that we have an "ideal window" and an "actual window", which is necessitated by the restrictions that apply to the actual window. For instance, when it is desired that the window contract, it may only do so one packet at a time by skipping a packet transmission when an acknowledgement is received. The ideal window, like rates, is not subject to such restrictions. The ideal window is adapted similarly to rates, while the actual window attempts to track the ideal window.

Let $K_j^{(I)}(t)$ denote the ideal window of VCj at time t. As in (2.1),

$$K_j^{(I)}(t) = \begin{cases} K_j^{(I)}(t - d_j(t)) - \gamma_j(u_j(t))d_j(t)[K_j^{(I)}(t - d_j(t)) - K_j^{\min}] + a_j(u_j(t))d_j(t)u_j(t) \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad j \in U(t) \\ K_j^{(I)}(t - 0) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if} \quad j \notin U(t) \end{cases} \tag{2.8}$$

Here $K_j^{\min}$ is related to the minimum rate ν_j; the specific relation is clarified in the next section. The functions $\gamma_j(\mu)$ and $a_j(\mu)$ are restricted as in (2.2).

Let $K_j(t)$ denote the actual window of VCj at time t. As mentioned above, a decrement of the actual window is necessarily of size 1. While increments may be larger, the description below is for the case where increments are also of size 1.

$$\begin{aligned} K_j(t) &= K_j(t-0) - 1 \quad \text{if} \quad K_j^{(I)}(t) \le K_j(t-0) - 1 \\ &= K_j(t-0) \qquad\quad \text{if} \quad K_j(t-0) - 1 < K_j^{(I)}(t) < K_j(t-0) + 1 \\ &= K_j(t-0) + 1 \quad \text{if} \quad K_j(t-0) + 1 \le K_j^{(I)}(t) \; . \end{aligned} \tag{2.9}$$

3 MODELS AND ANALYSIS FOR A SINGLE CONNECTION

In this section we concentrate on the case of just one virtual connection ($J = 1$), and prepare for the case of multiple connections, which is treated in the next section. A fluid model is introduced to approximate the algorithm described in Section 2 and certain basic properties of the model are established. In this section we suppress the virtual connection index j.

3.1 Model

The model for a single VC is

$$\frac{d}{dt}\,\phi(t) = \begin{cases} -\Gamma^+[\phi(t) - \nu] + A^+ u(t) & \text{if} \quad u(t) > 0 \\ -\Gamma^-[\phi(t) - \nu] + A^- u(t) & \text{if} \quad u(t) < 0 \end{cases} \tag{3.1}$$

$$\frac{d}{dt}\,Q(t) = \begin{cases} [\phi(t - \tau/2) - \mu] & \text{if} \quad Q(t) > 0 \\ [\phi(t - \tau/2) - \mu]^+ & \text{if} \quad Q(t) = 0 \end{cases} \tag{3.2}$$

and $u(t) = \text{sgn}\,[Q_T - Q(t - \tau/2)]$. In (3.2), $[x]^+ = \max(x, 0)$, and Γ^+, Γ^-, A^+, A^- are nonnegative parameters. Here $\phi(t)$ is the flow rate at the source transmitter and represents the fluid approximation to the local (in time) throughput of packets or cells. The classical queueing theoretic steady state throughput is approximated by the long-term average:

$\lim_{t\to\infty}\frac{1}{t}\int_0^t \phi(s)\,ds$. In (3.2) $Q(t)$ is an approximation to the queue in packets at the distant switch, and Q_T denotes the threshold. The nonlinear dynamics in (3.2) reflect the fact that the buffer content is nonnegative, i.e. $Q(t)\ge 0$. A similar restriction also applies to the rate $\phi(t)$, but this is not modelled. The reason is that $\phi(t)$ is typically large in our applications and the "floor" of $\phi(t)$ plays an insignificant role. This contrasts with our objective of keeping buffer contents small.

The correspondence with the description of the algorithm in (2.1) is obtained thus: $\phi(t)\approx R(t)$, and

$$\frac{d}{dt}\,\phi(t)\approx\frac{R(t)-R(t-d(t))}{d(t)}. \tag{3.3}$$

Hence the parameters of the models in (2.1) and (3.1) have simple correspondences:

$$\Gamma^+\approx\gamma^+,\quad \Gamma^-\approx\gamma^-,\quad A^+\approx a^+,\quad A^-\approx a^-. \tag{3.4}$$

The fluid model (3.1)–(3.2) may also be used to approximate the adaptive window algorithm, but the correspondences are somewhat different. The amount of fluid in the virtual connection at time t corresponding to a window $K^{(I)}(t)$ is approximately $\tau\phi(t)$. Proceeding as before, we let

$$\tau\,\frac{d}{dt}\,\phi(t)\approx\frac{K^{(I)}(t)-K^{(I)}(t-d(t))}{d(t)},$$

and, matching coefficients of (2.8) and (3.1), we obtain the following correspondences:

$$\Gamma^+\approx\gamma^+,\quad \Gamma^-\approx\gamma^-;\quad A^+\approx a^+/\tau,\quad A^-\approx a^-/\tau;\quad \nu\approx K^{\min}/\tau. \tag{3.5}$$

In particular, the important ratios $A^+/\Gamma^+=\frac{1}{\tau}(a^+/\gamma^+)$ and $A^-/\Gamma^-=\frac{1}{\tau}(a^-/\gamma^-)$.

Our design guidelines are first derived for the fluid models and then translated to the algorithms. Hence, a noteworthy point is that the design of algorithms for rate adaptation (see (3.4)) is independent of the propagation delay τ, while the algorithm for window adaptation (see (3.5)) depends on τ.

Equation (3.1) for the JRJ and MS algorithms has the following forms:

$$\text{JRJ:}\quad \frac{d}{dt}\,\phi(t) \;=\; Au(t)\quad\text{if}\quad u(t)>0 \tag{3.6}$$
$$\qquad\qquad\qquad =\; -\Gamma[\phi(t)-\nu]\quad\text{if}\quad u(t)<0$$

$$\text{MS:}\quad \frac{d}{dt}\,\phi(t) \;=\; -\Gamma[\phi(t)-\nu]+Au(t). \tag{3.7}$$

3.2 Regimes

The rate parameter ν in (2.1) and (3.1) is the user's expectation of minimum bandwidth. For this guarantee to hold it is necessary that admission control ensures

$$\nu \leq \rho\mu, \tag{3.8}$$

where ρ is the long-term channel utilization. In our investigations we have found ρ to be close to unity, more specifically in the range (0.80, 0.98).

Consider now the existence of stationary solutions, i.e., solutions of (3.1) and (3.2) with $\frac{d}{dt}\,\phi(t) = \frac{d}{dt}\,Q(t) = 0$. The only possibilities are

$$\phi = \nu + A^+/\Gamma^+, \quad \text{and} \quad \phi = \nu - A^-/\Gamma^-, \tag{3.9}$$

corresponding, respectively to $u \equiv +1$ and $u \equiv -1$. For the latter to hold, $Q > Q_T$ and $\phi > \mu$, and since $\mu > \nu$ from (3.8), it follows that $\phi > \nu$, which contradicts the expression in (3.9). Hence we may rule out $u \equiv -1$. The former stationary solution in (3.9) exists provided $(\nu + A^+/\Gamma^+) < \mu$. Hence we introduce the following

Definition. If $(\nu + A^+/\Gamma^+) < \mu$ then the system is Unsaturated, and if $(\nu + A^+/\Gamma^+) > \mu$ then the system is Saturated.

Proposition 3.1 (i) *If the system is Unsaturated then there exists a stationary solution* $\phi \equiv \nu + A^+/\Gamma^+$, $Q \equiv 0$.
(ii) *If the system is Saturated then no stationary solution exists.* ∎

Note that for the JRJ algorithm the system is always Saturated, since $\Gamma^+ = 0$. The following summarizes the dynamical behavior of Unsaturated systems.

Proposition 3.2 *Suppose the system is Unsaturated.*

(i) *For any t_1, if $\phi(t_1) \leq \nu + A^+/\Gamma^+$, then $\phi(t) \leq \nu + A^+/\Gamma^+$ for all $t \geq t_1$. If $\phi(t_0) > \nu + A^+/\Gamma^+$, then there exists t_1 ($t_1 > t_0$) such that $\phi(t)$ decreases monotonically in $[t_0, t_1)$ and $\phi(t_1) = \nu + A^+/\Gamma^+$.*

(ii) *There exists t_2 such that for all $t \geq t_2$,*

$$Q(t) < Q_T,$$

and

$$\phi(t) \to (\nu^+ + A^+/\Gamma^+)$$

at the exponential rate Γ^+.

Proof: (i) If $\phi(t) > \nu + A^+/\Gamma^+$ then from (3.1);

$$\frac{d}{dt}\,\psi(t) \leq \max[-\Gamma^+\psi(t) + A^+, \quad -\Gamma^-\psi(t) - A^-] < 0 \tag{3.10}$$

where,

$$\psi(t) \triangleq \phi(t) - \nu. \tag{3.11}$$

(ii) Say $\phi(t) \leq \nu + A^+/\Gamma^+$ for all $t \geq t_1$. Since the system is Unsaturated

$$\phi(t - \tau/2) - \mu < 0 \qquad (t \geq t_1 + \tau/2)$$

and consequently, from (3.2), for all $t \geq t_1 + \tau/2$,

$$\begin{aligned} \frac{d}{dt}\,Q(t) &< 0 \quad \text{if} \quad Q(t) > 0 \\ &= 0 \quad \text{if} \quad Q(t) = 0. \end{aligned} \tag{3.12}$$

Hence after $(t_1 + \tau/2)$ the buffer content decreases monotonically till the buffer is empty and thereafter it remains empty. Hence, in particular, there exists t_2 such that for all $t \geq t_2$,

$$Q(t) < Q_T \quad \text{and} \quad u(t) \equiv 1.$$

Hence, from (3.1), for all $t \geq t_2$,

$$\phi(t) = [\nu + A^+/\Gamma^+] + e^{-\Gamma^+(t-t_2)}[\phi(t_2) - \nu - A^+/\Gamma^+]. \tag{3.13}$$

Since $\Gamma^+ > 0$ for Unsaturated systems, the result follows. ∎

Observe that the queue threshold Q_T does not have a role in determining the steady state values of $Q(t)$ and $\phi(t)$.

We remark that for the symmetric MS algorithm it is possible to sharpen the statements made above. For instance, it follows from (3.10) that for this case,

$$\frac{d}{dt}\,\psi(t) \leq -\Gamma\psi(t) + A, \tag{3.14}$$

and hence the approach to A/Γ from above is exponentially fast.

We now give a basic boundedness result which holds for systems, possibly asymmetric, which are "uniformly damped", i.e.,

$$\Gamma_M \triangleq \min(\Gamma^+, \Gamma^-) > 0. \tag{3.15}$$

Note that this result holds regardless of whether the system is Saturated or Unsaturated.

Proposition 3.3 *Suppose the system is uniformly damped and let $A_M \triangleq \max(A^+, A^-)$. Then for any t_0 and $t > t_0$,*

$$|\phi(t) - \nu| \leq A_M/\Gamma_M + e^{-\Gamma_M(t-t_0)}[|\phi(t_0) - \nu| - A_M/\Gamma_M]. \tag{3.16}$$

Proof: With $\psi(t)$ defined in (3.11) and $\frac{d^+}{dt}\,|\psi(t)|$ defined as the right derivative [COP65],

$$\begin{aligned}
\frac{d^+}{dt}\,|\psi(t)| &= [\text{sgn }\psi(t)]\frac{d}{dt}\,\psi(t) \\
&= \begin{cases} [\text{sgn }\psi(t)]\,[-\Gamma^+\psi(t) + A^+] & \text{if} \quad u(t) > 0 \\ [\text{sgn }\psi(t)]\,[-\Gamma^-\psi(t) - A^-] & \text{if} \quad u(t) < 0 \end{cases} \\
&= \begin{cases} -\Gamma^+|\psi(t)| + A^+[\text{sgn }\psi(t)] & \text{if} \quad u(t) > 0 \\ -\Gamma^-|\psi(t)| - A^-[\text{sgn }\psi(t)] & \text{if} \quad u(t) < 0 \end{cases} \\
&\leq -\Gamma_M|\psi(t)| + A_M.
\end{aligned} \tag{3.17}$$

Hence, for $t > t_0$,

$$|\psi(t)| \leq A_M/\Gamma_M + e^{-\Gamma_M(t-t_0)}[|\psi(t_0)| - A_M/\Gamma_M]. \tag{3.18}$$

■

3.3 Periodic Solutions

We proceed to the Saturated regime where it is known from Proposition 3.1 that no stationary solutions exist. We show (see Figure 3.1) by construction that periodic solutions (limit cycles) exist in this regime. To keep it simple, the construction is for the MS algorithm with $\nu = 0$. The technique may be extended to cover other cases, albeit at the cost of introducing various tedious details. This construction is related to one in [FRW92], with the main difference being the presence of damping here; also, [FRW92] considered various nonlinearities. See also [MST91] for construction of limit cycles for the JRJ algorithm.

Start by postulating a point t_0 such that for $t \in [t_0 - \tau/2,\ t_0]$,

$$\begin{aligned}
&\phi(t) = \frac{A}{\Gamma}\left[1 - e^{-\Gamma(t-t_0+\tau/2)}\right] + \mu e^{-\Gamma(t-t_0+\tau/2)} \\
&Q(t) = 0
\end{aligned} \tag{3.19}$$

Note that the postulated $\phi(t)$ is a solution of the equations

$$\left.\begin{aligned}
&\frac{d}{dt}\,\phi(t) = -\Gamma\phi(t) + A \\
&\phi(t_0 - \tau/2) = \mu
\end{aligned}\right\} \tag{3.20}$$

We will show that at a later point, t_3,

$$\phi(t_3 - s) = \phi(t_0 - s), \quad \text{and} \quad Q(t_3 - s) = Q(t_0 - s) \quad (0 \leq s \leq \tau/2) \tag{3.21}$$

This will establish the existence of a periodic solution to (3.1)–(3.2). There are three phases present in the limit cycle. In Phase 1, $t \in (t_0, t_1)$; in Phase 2, $t \in (t_1, t_2)$; in Phase 3, $t \in (t_2, t_3)$, where t_1, t_2 and t_3 are explicitly identified in the construction given below.

The key feature in Phase 1 is that

$$Q(t - \tau/2) < Q_T \qquad (t \in \text{ Phase 1}) \tag{3.22}$$

Phase 1 terminates at t_1 where

$$Q(t_1 - \tau/2) = Q_T. \tag{3.23}$$

Hence, from (3.1) and (3.2), $u(t) = 1$ for $t \in$ Phase 1. Hence,

$$\frac{d}{dt}\,\phi(t) = -\Gamma\phi(t) + A \qquad (t \in \text{Phase 1}). \tag{3.24}$$

Now, since the system is Saturated it follows from (3.23) that $\phi(t_0) > \mu$, and hence,

$$\phi(t) > \mu, \;\; \frac{d}{dt}\,\phi(t) > 0 \qquad (t \in \text{Phase 1}). \tag{3.25}$$

Hence,

$$\frac{d}{dt}\,Q(t) = \phi(t - \tau/2) - \mu > 0 \qquad (t \in \text{Phase 1}). \tag{3.26}$$

The time t_1, see (3.23), is calculated from (3.24) and (3.26). As shown in Figure 3.1, $Q(t_1)$ overshoots Q_T.

Phase 2 terminates at t_2 where

$$Q(t_2 - \tau/2) = Q_T \tag{3.27}$$

and,

$$Q_T < Q(t - \tau/2) \qquad (t \in \text{ Phase 2}). \tag{3.28}$$

Hence, the governing equations for Phase 2 are

$$\begin{aligned} \frac{d}{dt}\,\phi(t) &= -\Gamma\phi(t) - A \\ \frac{d}{dt}\,Q(t) &= \phi(t - \tau/2) - \mu \end{aligned} \tag{3.29}$$

Note that $d\phi/dt < 0$. Hence $Q(t)$ is concave and reaches a maximum value $\hat{Q}$ at time denoted by t_2^*. Also, $\phi(t_2) < \mu$.

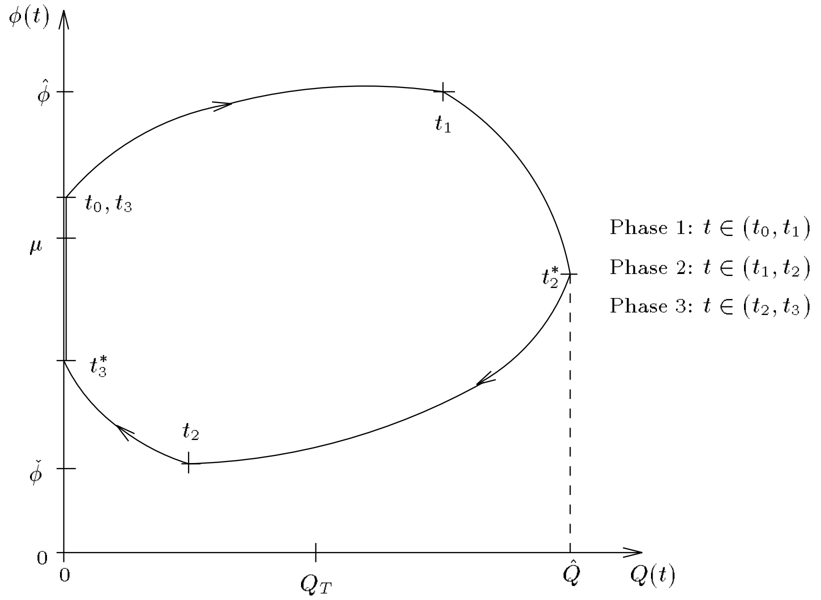

Figure 3.1: Sketch of periodic solution in Saturated regime.

In Phase 3,

$$\phi(t - \tau/2) < \mu \quad \text{and} \quad Q(t - \tau/2) < Q_T \tag{3.30}$$

The governing equations are

$$\begin{aligned} \frac{d}{dt}\,\phi(t) &= -\Gamma\phi(t) + A \\ \frac{d}{dt}\,Q(t) &= \phi(t - \tau/2) - \mu \qquad && t_2 < t < t_3^* \\ &= 0 && t > t_3^* \end{aligned} \tag{3.31}$$

where t_3^* is the time when the buffer first empties. Note that $d\phi/dt > 0$ and $dQ(t)/dt \leq 0$. It is not hard to show that $\phi(t_3^* - \tau/2) < \mu$. Hence, there exists $t_3 > t_3^*$ such that $\phi(t_3 - \tau/2) = \mu$, which is the defining relation for t_3, the terminal time point of Phase 3 and the start of Phase 1. To see that the solution is periodic, i.e. (3.21) holds, it suffices to observe that the differential equations governing $\phi(t)$ in (3.20) and (3.31) are identical.

Note that the maximum value of $\phi(t)$, $\hat{\phi}$, is reached at $t = t_1$, and the minimum value, $\check{\phi}$, is reached at $t = t_2$.

A limiting case of the Saturated Regime is where the damping, $\Gamma = 0$ and $A/\Gamma = \infty$. It is particularly interesting that even in this case the limit cycle exists and is bounded. As it happens, the construction is simpler with $\Gamma = 0$ and the following explicit expressions are obtained for the extremal quantities, $\hat{Q}$, $\hat{\phi}$ and $\check{\phi}$:

$$\hat{Q} = \frac{A}{2}\left[\left(\frac{\tau}{2}+\sqrt{\frac{2Q_T}{A}}\right)^2+\left(\frac{3\tau}{2}+\sqrt{\frac{2Q_T}{A}}\right)^2\right] \tag{3.32}$$

$$\hat{\phi} = \mu + A\tau + \sqrt{2Q_T A} \tag{3.33}$$

$$\check{\phi} = \mu - A\tau - \left[\left(2A\tau+\sqrt{2AQ_T}\right)^2-\frac{3}{2}(A\tau)^2\right]^{1/2} \tag{3.34}$$

4 MULTIPLE CONNECTIONS: FAIRNESS AND DESIGN

This section breaks from the preceding section in its focus on multiple connections and the Saturated operating regime. Fairness is a dominating theme with multiple connections and most of this section is concerned with, first, defining and then achieving, through design of algorithms, various forms of fair behavior. The Saturated operating regime is the one of choice, even at the cost of oscillatory system behavior, since capacity is more completely utilized. Finally, the topic of interoperability, i.e., the co-existence of diverse algorithms, is investigated.

4.1 Fairness

Each VC has two associated parameters for purposes of bandwidth allocation, (ν_j, σ_j) for VCj. Where ϕ_j is the bandwidth realized by VCj, it is desired that

$$\phi_j \geq \nu_j, \quad \text{for all} \quad j, \tag{4.1}$$

and

$$\frac{\phi_j-\nu_j}{\sigma_j} = \frac{\phi_{j'}-\nu_{j'}}{\sigma_{j'}}, \quad \text{for all pairs} \quad (j,j'). \tag{4.2}$$

Note that $\{\sigma_j\}$ are relative weights and hence are not required to sum to unity. It is tacitly understood that admission control is administered to enforce the condition

$$\sum_j \nu_j < \rho\mu, \tag{4.3}$$

where μ is the network bandwidth, and ρ is the utilization. Admission control is not necessary if all $\nu_j = 0$.

Service classes of particular interest are,

(i) Best Effort: $\nu_j = 0$, $\sigma_j = \sigma > 0$.

(ii) Best Effort with Minimum Bandwidth: $\nu_j > 0$, $\sigma_j = \sigma > 0$.

Other services may similarly be defined by appropriate selection of the parameters ν_j and σ_j. Typically service will be mixed, i.e., service classes will coexist with each class comprising of several VCs.

4.2 Model

We extend the model in (3.1) and (3.2) in the natural way. Let $(\Gamma_j^+, A_j^+, \Gamma_j^-, A_j^-)$ be parameters associated with VCj in this model. For $j = 1, 2, \ldots, J$,

$$\frac{d}{dt}\,\phi_j(t) = \begin{cases} -\Gamma_j^+[\phi_j(t) - \nu_j] + A_j^+ u_j(t) & \text{if} \quad u_j(t) > 0 \\ -\Gamma_j^-[\phi_j(t) - \nu_j] + A_j^- u_j(t) & \text{if} \quad u_j(t) < 0 \end{cases} \tag{4.4}$$

$$\frac{d}{dt}\,Q(t) = \begin{cases} \left[\sum_j \phi_j(t - \tau_j/2) - \mu\right] & \text{if} \quad Q(t) > 0 \\ \left[\sum_j \phi_j(t - \tau_j/2) - \mu\right]^+ & \text{if} \quad Q(t) = 0 \end{cases} \tag{4.5}$$

and $u_j(t) = \text{sgn}\,[Q_T - Q(t - \tau_j/2)]$. For the JRJ algorithm $\Gamma_j^+ = A_j^- = 0$ for all j; let $\Gamma_j = \Gamma_j^-$ and $A_j = A_j^+$. For the symmetric MS algorithm $\Gamma_j^+ = \Gamma_j^- = \Gamma_j$ and $A_j^+ = A_j^- = A_j$.

The system is defined to be Saturated or Unsaturated depending upon whether $\sum_j(\nu_j + A_j^+/\Gamma_j^+)$ is greater or less than μ, respectively. The implication of the regime on the existence of stationary solutions, which is stated for a single VC in Proposition 3.1, also holds for multiple connection. Proposition 3.2 extends naturally and states that the dynamical behavior is such that the stationary solution is approached from all initial conditions if the system is Unsaturated. Finally, dynamical behavior is oscillatory if the system is Saturated, but bounded, as stated in Proposition 3.3.

In the Unsaturated regime the unutilized bandwidth is $[\mu - \sum_j(\nu_j + A_j^+/\Gamma_j^+)] > 0$. This is a strong incentive to operate in the Saturated regime. Also, it is typically inevitable that systems will sometimes operate in the Saturated regime. This is because, in long time scales, the number of virtual circuits, J, will vary substantially and, even if the design is Unsaturated for small J, it will typically be Saturated at the other extreme of large J.

4.3 Pointwise Fairing

We show here that with proper design, the divergences from fair bandwidth allocation in all our algorithms vanish monotonically, even exponentially in certain algorithms. Define

$$\theta_j(t) \triangleq \phi_j(t + \tau_j/2) - \nu_j \qquad (j = 1, 2, \ldots, J). \tag{4.6}$$

Then,

$$\frac{d}{dt}\,\theta_j(t) = \begin{cases} -\Gamma_j^+\theta_j(t) + A_j^+u(t) & \text{if} \quad u(t) > 0 \\ -\Gamma_j^-\theta_j(t) + A_j^-u(t) & \text{if} \quad u(t) < 0 \end{cases} \tag{4.7}$$

where $u(t) = \text{sgn}\,[Q_T - Q(t)]$. Note that with this change of variables the forcing function $u(t)$ is common to all j in (4.7). Also,

$$u(t) = \begin{cases} \dfrac{d}{dt}\left(\dfrac{\theta_j}{A_j^+}\right) + \left(\dfrac{\Gamma_j^+}{A_j^+}\right)\theta_j & \text{if} \quad u(t) > 0 \\ \dfrac{d}{dt}\left(\dfrac{\theta_j}{A_j^-}\right) + \left(\dfrac{\Gamma_j^-}{A_j^-}\right)\theta_j & \text{if} \quad u(t) < 0. \end{cases} \tag{4.8}$$

We propose the following **design rule**: for $j = 1, 2, \ldots, J$

$$\begin{aligned} A_j^+ &= A^+\sigma_j & \qquad \Gamma_j^+ &= \Gamma^+ \\ A_j^- &= A^-\sigma_j & \qquad \Gamma_j^- &= \Gamma^- \end{aligned} \tag{4.9}$$

where $A^+, A^-, \Gamma^+, \Gamma^-$ are nonnegative constants common to all VCs.

From (4.8) and (4.9), for any pair j, j',

$$\frac{d}{dt}\left[\frac{\theta_j}{\sigma_j} - \frac{\theta_{j'}}{\sigma_{j'}}\right] = \begin{cases} -\Gamma^+\left[\dfrac{\theta_j}{\sigma_j} - \dfrac{\theta_{j'}}{\sigma_{j'}}\right] & \text{if} \quad u(t) > 0 \\ -\Gamma^-\left[\dfrac{\theta_j}{\sigma_j} - \dfrac{\theta_{j'}}{\sigma_{j'}}\right] & \text{if} \quad u(t) < 0. \end{cases} \tag{4.10}$$

Define $\nabla_{jj'}(t)$ to be the "divergence from fairness",

$$\begin{aligned} \nabla_{jj'}(t) &\triangleq \frac{\phi_j(t+\tau_j/2) - \nu_j}{\sigma_j} - \frac{\phi_{j'}(t+\tau_{j'}/2) - \nu_{j'}}{\sigma_{j'}} \\ &= \frac{\theta_j(t)}{\sigma_j} - \frac{\theta_{j'}(t)}{\sigma_{j'}}. \end{aligned} \tag{4.11}$$

From (4.10),

$$\frac{d}{dt}\,\nabla_{jj'}(t) = -c(t)\nabla_{jj'}(t), \tag{4.12}$$

where

$$c(t) \triangleq \Gamma^+ I_{u(t)>0} + \Gamma^- I_{u(t)<0}. \tag{4.13}$$

Note that $c(t) \geq 0$, for all t. Now, from (4.12), for any $t_0 < t$,

$$\nabla_{jj'}(t) = \nabla_{jj'}(t_0)\exp\left[-\int_{t_0}^{t} c(s)ds\right]. \tag{4.14}$$

If the system is Unsaturated, then $\nabla_{jj'}(t) \to 0$ as $t \to \infty$, since, as previously noted, $\phi_j(t) \to \nu_j + A_j^+/\Gamma_j^+ = \nu_j + \sigma_j A^+/\Gamma^+$. If the system is Saturated, then the expansionary and contractive phases recur infinitely often and consequently (see (4.13)), $\int_{t_0}^t c(s)ds \to \infty$ as $t \to \infty$. Hence, from (4.14), in this case too, $\nabla_{jj'}(t) \to 0$ as $t \to \infty$.

This property, wherein the divergences from fair bandwidth allocation vanish monotonically, is what we mean by "pointwise fairing".

Thus far the focus has been on the comparative behavior of VCs. Insight into the aggregate behavior is obtained by summing (4.7) over all virtual circuits. Let

$$\theta(t) \triangleq \sum_{j=1}^{J} \theta_j(t). \tag{4.15}$$

Then, from the design rule (4.9),

$$\frac{d}{dt}\,\theta(t) = \begin{cases} -\Gamma^+\theta(t) + A^+(\sum_j \sigma_j)u(t) & \text{if} \quad u(t) > 0 \\ -\Gamma^-\theta(t) + A^-(\sum_j \sigma_j)u(t) & \text{if} \quad u(t) < 0. \end{cases} \tag{4.16}$$

That is, $\theta(t)$ behaves like the bandwidth of a single VC for which the algorithm parameters are Γ^+, $A^+(\sum_j \sigma_j)$, Γ^-, $A^-(\sum_j \sigma_j)$. The analysis of Section 3 may be brought to bear on this quantity.

Recall that our prior results in this section have shown that $\theta_j(t) \to \sigma_j C(t)$, for all j, where $C(t)$ is a common function. This fact gives

$$\theta_j(t) \to \frac{\sigma_j}{(\sum_k \sigma_k)}\,\theta(t) \quad \text{as} \quad t \to \infty. \tag{4.17}$$

The following is a summary of our results, which hold regardless of whether the system is Saturated or Unsaturated.

Proposition 4.1 *Let the parameters of the algorithm be chosen by the design rule* (4.9), *and let* $\nabla_{jj'}(t)$ *in* (4.11) *be a measure of the divergence from fair allocation for any two VCs,* j *and* j'.

(i) *For all initial conditions,* $\nabla_{jj'}(t)$ *is monotonic, nonincreasing and approaches* 0 *as* $t \to \infty$.

(ii) *Let* $\theta(t)$ *in* (4.15) *denote the aggregate bandwidth.* (4.17) *gives the asymptotic behaviors of bandwidths of individual VCs in terms of the aggregate.* $\theta(t)$ *is obtained as the solution of the model for a single VC.* ■

Let us now examine in somewhat greater detail the implications of the design rule on the JRJ and MS algorithms. For the JRJ algorithm, from (4.13),

$$\begin{aligned} c(t) &= 0 \quad \text{if} \quad u(t) > 0 \\ &= \Gamma \quad \text{if} \quad u(t) < 0. \end{aligned} \tag{4.18}$$

That is, the divergence from fairness remains constant during the expansionary phases and decreases exponentially during the contractive phases.

In contrast, for the symmetric MS algorithm, for all t,

$$c(t) \equiv \Gamma, \tag{4.19}$$

and consequently,

$$\nabla_{jj'}(t) = \nabla_{jj'}(t_0) e^{-\Gamma(t-t_0)}, \tag{4.20}$$

i.e., the approach to fairness is exponentially fast with rate Γ. This is an important and attractive feature of the algorithm.

4.4 Averaging Based on Periodicity

The results of this subsection apply only to the MS algorithm in the Saturated regime. Specifically, symmetry in the general form of the algorithm is used in an essential manner; also, the results here depend on the periodicity of $\{\phi_j(t)\}$ and $Q(t)$, which are the observed forms of behavior of the Saturated system. The benefit is a relaxation of the design rules in (4.9), which is of considerable convenience in the design for VCs with very different geographical lengths and updating frequencies. In return we give up exponential pointwise convergence to fair bandwidth allocation for "averaged fairness", i.e., the fair allocation of the time-averaged bandwidths of the VCs.

Specializing (4.4) to the MS algorithm, we obtain

$$\frac{d}{dt}\,\psi_j(t) = -\Gamma_j \psi_j(t) + A_j u_j(t) \quad (j = 1, 2, \ldots, J) \tag{4.21}$$

where $\psi_j(t) \triangleq \phi_j(t) - \nu_j$. Consider continuous periodic solutions of (4.21) with period T, i.e., for all j and t,

$$\psi_j(t) = \psi_j(t+T), \quad \text{and} \quad u_j(t) = u_j(t+T). \tag{4.22}$$

Integrating (4.21) over the interval $[s, s+T]$ and noting that $\int_s^{s+T} \frac{d}{dt}\,\psi_j(t)dt = 0$, we obtain

$$\int_s^{s+T} \psi_j(t)dt = (A_j/\Gamma_j) \int_s^{s+T} u_j(t)dt. \tag{4.23}$$

Define the time averages,

$$\langle \phi_j \rangle \triangleq \frac{1}{T} \int_s^{s+T} \phi_j(t)dt, \quad \langle u \rangle \triangleq \frac{1}{T} \int_s^{s+T} u_j(t)dt. \tag{4.24}$$

Then from (4.23),

$$\frac{\langle \phi_j \rangle - \nu_j}{(A_j/\Gamma_j)} = \langle u \rangle \qquad (j = 1, 2, \ldots, J) \tag{4.25}$$

which is an important result.

Let fair behavior in the averaged sense refer to the following condition for any pair (j, j'):

$$\frac{\langle \phi_j \rangle - \nu_j}{\sigma_j} = \frac{\langle \phi_{j'} \rangle - \nu_{j'}}{\sigma_{j'}}. \tag{4.26}$$

From (4.25) we see that fairness in the averaged sense is achieved by the MS algorithm provided the following design rule is followed:

$$A_j/\Gamma_j = (\text{constant})\ \sigma_j \quad (j = 1, 2, \ldots, J) \tag{4.27}$$

where $A_j \geq 0$, $\Gamma_j \geq 0$. This is a considerable relaxation over the design rule in (4.9). Of course, if (4.9) is satisfied, then so is (4.27).

We note in passing that if the system is Unsaturated then such a design rule gives the desired behavior in (4.26), provided $\langle \phi_j \rangle$ is interpreted as the stationary value of ϕ_j, for each j.

A necessary condition for periodic behavior is that the system is Saturated. In this case

$$\sum_j \langle \phi_j \rangle \approx \rho\mu. \tag{4.28}$$

Combining (4.25) and (4.28),

$$\langle u \rangle = \left(\rho\mu - \sum_j \nu_j \right) / \sum_j (A_j/\Gamma_j). \tag{4.29}$$

Hence if admission control is administered, i.e., $\sum \nu_j < \rho\mu$, then $\langle u \rangle > 0$, and consequently, from (4.25),

$$\langle \phi_j \rangle > \nu_j, \tag{4.30}$$

i.e., the soft version of the guarantee on minimum bandwidths is satisfied.

4.5 Averaging Based on Small Amplitude Oscillations

When we consider algorithms, like the JRJ algorithm, which behave asymmetrically in its expansionary and contractive phases, the arguments in the preceding subsection leading to averaging based on just periodicity break down, as we shall see. To cope with this situation we make the more restrictive assumption of small amplitude oscillations to arrive at equations for the time averages of rates. These equations also apply to the symmetric MS algorithm, and it is reassuring that there is agreement between the result thus obtained and the prior result based on periodicity. The assumption made here states that the oscillations in the rates have high frequency, and hence may equivalently be viewed as a form of time scale separation.

From (4.4),

$$\frac{d}{dt}\,\psi_j(t) = \begin{cases} -\Gamma_j^+\psi_j(t) + A_j^+ u_j(t) & \text{if} \quad u_j(t) > 0 \\ -\Gamma_j^-\psi_j(t) + A_j^- u_j(t) & \text{if} \quad u_j(t) < 0 \end{cases} \tag{4.31}$$

where $\psi_j(t) = \phi_j(t) - \nu_j$. Assume, as in (4.22), that a periodic solution of period T exists. Let

$$T^+ \triangleq \int_s^{s+T} I_{u_j(t)>0}\,dt, \quad T^- \triangleq \int_s^{s+T} I_{u_j(t)<0}\,dt \tag{4.32}$$

and let p denote the fraction of time that the feedback signal is positive, i.e.,

$$p = \frac{T^+}{T^+ + T^-}. \tag{4.33}$$

Integrating (4.31) over $[s, s+T]$ gives

$$\begin{aligned}\Gamma_j^+ \int_s^{s+T} \psi_j(t) I_{u_j(t)>0}\,dt \; - \; A_j^+ \int_s^{s+T} I_{u_j(t)>0}\,dt + \Gamma_j^- \int_s^{s+T} \psi_j(t) I_{u_j(t)<0}\,dt \\ + \; A_j^- \int_s^{s+T} I_{u_j(t)<0}\,dt = 0,\end{aligned} \tag{4.34}$$

since, from periodicity, $\int_s^{s+T} \frac{d}{dt}\,\psi_j(t)\,dt = 0$. Let

$$\langle \psi_j^+ \rangle \triangleq \frac{1}{T^+} \int_s^{s+T} \psi_j(t) I_{u_j(t)>0}\,dt, \quad \text{and} \quad \langle \psi_j^- \rangle \triangleq \frac{1}{T^-} \int_s^{s+T} \psi_j(t) I_{u_j(t)<0}\,dt. \tag{4.35}$$

From (4.34) and (4.35),

$$p[\Gamma_j^+\langle \psi_j^+ \rangle - A_j^+] + (1-p)[\Gamma_j^-\langle \psi_j^- \rangle + A_j^-] = 0. \tag{4.36}$$

In the asymmetric case this line of reasoning does not take us any further.

It is instructive at this point to examine the JRJ algorithm for a single VC.

$$\begin{aligned} \frac{d}{dt}\,\psi(t) &= Au(t) \quad \text{if} \quad u(t) > 0 \\ &= -\Gamma\psi(t) \quad \text{if} \quad u(t) < 0. \end{aligned} \tag{4.37}$$

Typical solutions are sketched below

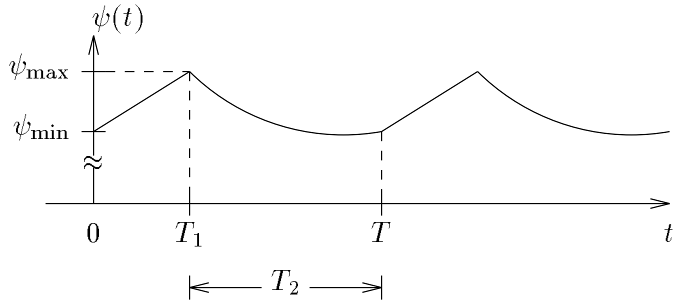

Figure 4.1:

Solving (4.37), we get

$$\begin{aligned} \psi_{\text{max}} &= \psi_{\text{min}} + AT_1 \\ \psi_{\text{min}} &= \psi_{\text{max}} e^{-\Gamma T_2}. \end{aligned} \tag{4.38}$$

Hence,

$$\frac{\psi_{\text{max}} - \psi_{\text{min}}}{\psi_{\text{max}}} = 1 - e^{-\Gamma T_2}. \tag{4.39}$$

The oscillations are of small amplitude, i.e., $(\psi_{\text{max}} - \psi_{\text{min}})/\psi_{\text{max}} \ll 1$, if and only if ΓT_2 is small.

We proceed by returning to (4.36), which was obtained from periodicity, and make the additional assumption of small amplitude oscillations. Then, see (4.35),

$$\langle \psi_j^+ \rangle \approx \langle \psi_j^- \rangle. \tag{4.40}$$

Let $\langle \psi_j \rangle$ denote the (approximately) common quantity in (4.40). Substituting in (4.36) we obtain for the general algorithm,

$$\langle \psi_j \rangle = \frac{pA_j^+ - (1-p)A_j^-}{p\Gamma_j^+ + (1-p)\Gamma_j^-} \qquad (j = 1, 2, \ldots, J). \tag{4.41}$$

In the special case of the JRJ algorithm (4.41) yields

$$\langle \psi_j \rangle = \frac{p}{1-p}\,\frac{A_j}{\Gamma_j}, \tag{4.42}$$

and for the MS algorithm,

$$\langle \psi_j \rangle = (2p-1)\frac{A_j}{\Gamma_j}. \tag{4.43}$$

On noting that $(2p-1) = \langle u \rangle$, we see that (4.43) is equivalent to (4.25). This is reassuring because (4.25) was obtained from substantially weaker assumptions.

We may make use of (4.42) to design the JRJ algorithm for fair behavior in the averaged sense, which is described by (4.26). The design rule is simply (4.27).

Note that since the right hand quantity in (4.42) is positive, it follows immediately that

$$\langle \psi_j \rangle = \langle \phi_j \rangle - \nu_j > 0 \tag{4.44}$$

for all j, i.e., the soft version of the guarantee on minimum bandwidth is satisfied by the JRJ algorithm.

We end this subsection by reminding the reader that the results of this subsection are based on the assumption of small amplitude oscillations, which limits the applicability of the results. However, further inferences are drawn from this model in the following subsections and checked with simulations. The results are reported later. It suffices to mention here that the results support the usefulness of the analysis.

4.6 Discussion of Design Procedures

Here we recapitulate and discuss briefly the implications of the preceding analyses on design and operating procedures. Our approach is to do our primary design for the continuous-time fluid models and subsequently translate back to the parameters of the actual algorithm in (2.1) and (2.5). The advantage is that the fluid models are more tractable and provide a uniform framework. The disadvantage is that the translation can introduce undesirable artifacts when the time intervals between updates in the algorithm are excessively lengthy. However, the latter should be avoided in any case. The translation we employ is straightforward and given by (3.4) for rates and by (3.5) for windows.

An important feature of the algorithms discussed here is the extraordinary robustness of performance over a broad range of parameters corresponding to various degrees of saturatedness. This feature will be amply exhibited in the simulation results to be presented in the next section. To be specific, let us consider design for pointwise fairing and refer to (4.16), which gives the governing equation for the aggregate bandwidth $\theta(t)$. Consider Γ^+, A^+, Γ^-, and A^- to be fixed, while J and therefore $(\sum_{j=1}^{J} \sigma_j)$ are varied. As the latter quantity, which, say, is denoted by Σ_J, is increased, the system becomes increasingly Saturated with increasing energy in the fluctuations of $\theta(t)$ and $Q(t)$, the buffer constant. However, the growth in the fluctuations is surprisingly slow and over an order of magnitude change in Σ_J the realized bandwidth remains close to the network bandwidth, μ.

The aforementioned feature is, of course, central to the adaptability of the algorithm, by which we mean robust behavior over a wide range of the number of VCs, J. This is essential since J and Σ_J vary with time, and when new connections are set up or old connections torn down, the other connections' parameters are not modified. Such adaptability may be thought of as single scale adaptability. The robustness discussed above breaks down when Σ_J is varied over several orders of magnitude. To cope with this situation, i.e., to extract multiple scale adaptability, it is necessary to rescale parameters. For example, we may consider scales in which $1 \leq J \leq 20$ and $21 \leq J \leq 200$, and the dwell time in each scale to be fairly long. In this case, we propose that the parameters of all VCs are rescaled at the time of a change of scale.

4.7 Interoperability

So far we have covered systems in which all VCs operate according to a common algorithm. What if some VCs, say in group 1, operate according to the JRJ algorithm and other VCs, say in group 2, operate according to the MS algorithm? Our prior analysis allows us to predict steady state behavior in such a configuration. First, all VCs in each group behave fairly, in accordance to the design rule employed. However, by contrasting (4.42) and (4.43) it is clear that representative behavior of the two groups will be different. For instance, if all VCs have a common value of A_j/Γ_j, then since $\{p/(1-p)\} > (2p-1)$ for all p in $(0,1)$, it follows that the bandwidth obtained by VCs in group 1 will be more than that of VCs in group 2. This should not be surprising since we have seen (recall (4.18) and (4.19), for instance) that at a detailed level the algorithms share bandwidth differently. The implication, however, is that interoperability of algorithms is very limited.

Here we show that interoperability of the JRJ and MS algorithms can be designed for certain limited frameworks. Suppose that VCs $1, 2, \ldots, N_1$ comprise group 1 and VCs $(N_1+1), \ldots, (N_1+N_2)$ comprise group 2. Further suppose that the goal is to provide Best Effort service, i.e., $\nu_j \equiv 0$ and $\sigma_j \equiv \sigma$ for $1 \leq j \leq (N_1+N_2)$. Accordingly let

$$\begin{aligned} A_j/\Gamma_j &= A_1/\Gamma_1 \quad j = 1, 2, \ldots, N_1 \\ &= A_2/\Gamma_2 \quad j = (N_1+1), (N_1+2), \ldots, (N_1+N_2). \end{aligned} \tag{4.45}$$

Now, from (4.42) and (4.43), we require

$$\frac{p}{1-p}\frac{A_1}{\Gamma_1} = (2p-1)\frac{A_2}{\Gamma_2} \tag{4.46}$$

and, from (4.28),

$$N_1\frac{p}{1-p}\frac{A_1}{\Gamma_1} + N_2(2p-1)\frac{A_2}{\Gamma_2} = \rho\mu. \tag{4.47}$$

For given N_1 and N_2, we may solve (4.46) and (4.47) for $(A_1/\Gamma_1)/(A_2/\Gamma_2)$. Let

$$\phi^* \triangleq \frac{\rho\mu}{N_1 + N_2} \tag{4.48}$$

denote the fair share of bandwidth for each VC. Then,

$$\frac{A_1}{\Gamma_1} = \phi^* \ \frac{A_2/\Gamma_2 - \phi^*}{A_2/\Gamma_2 + \phi^*}. \tag{4.49}$$

Hence, if the above design rule is followed then in steady state all VCs will receive a fair share of bandwidth. Observe that as expected, in this design

$$\frac{A_1}{\Gamma_1} < \phi^* < \frac{A_2}{\Gamma_2}. \tag{4.50}$$

Simulation results confirm that the algorithm design in (4.49) gives fair behavior.

5 SIMULATIONS

Extensive simulations of the algorithms in Section 2 were done in order to illustrate, qualitatively and quantitatively, some of the issues discussed above. The simulations used PANACEA [RMI88], a software package for the performance of queueing networks. The adaptive controls, which are nonstandard in the queueing network framework, were implemented by exploiting the "user-escape" mechanism designed for nonstandard applications [SEE88]. The simulated entities are ATM cells.

In this presentation we limit our consideration to:

- Rate-based flow control. Window-based implementations have also been simulated, but are not described here.

- Infinite data sources, i.e., once it turns on, a VC continues indefinitely to generate traffic at the controlled rate.

- Periodic rate updating, following the synchronous model of (2.5) in Section 2. The updating interval ranges from 0.5τ to 2.0τ.

- Single point of congestion, or queue, characterized by the FCFS service discipline; buffer overflows are not modelled.

- Fixed packet size of 53 bytes, the standard ATM cell size, and transmission rates of either 155 Mbps or 45 Mbps. In most cases the round-trip propagation delay is 10 msec, and in one case it is 100 msec.

The issues explored in the following subsections include: (i) the general behavior of the algorithms as a function of their key parameters (e.g., gain, damping, queue threshold, round-trip delays, rate updating intervals); (ii) fairness for multiple virtual connections; (iii) interoperability; (iv) adaptability as a function of the number of active virtual connections.

5.1 Effect of Gain, Damping and Other Relevant Parameters

In this section we investigate the effect of the gain parameter, a, and the damping constant, γ, on both the steady state and the transient behavior of the source rates.

Figure 5.1 considers a single VC ($J = 1$) in which the source rate is adapted by the MS algorithm. The damping constant γ is fixed while the gain a is varied. It is easy to identify cases in the Unsaturated and the Saturated regimes. Note that rates only are plotted; queues are not shown. The overall conclusion is that it is desirable to operate in the Saturated regime, where the rise time is smaller, the steady state throughput is maximal and the oscillations are modest. Even in the extreme example of the top curve, where $a/\gamma + \nu \approx 3.7\mu$, the behavior is acceptable.

A major issue investigated in Figures 5.2–5.3 is whether the algorithms are fair; i.e., whether a second VC, which turns on when the first VC is already using most or all of the network bandwidth, succeeds in ultimately obtaining its share of the bandwidth, assumed here to be half. In Figures 5.2 and 5.3, $J = 1$ for $t < 2.5$ sec and $J = 2$ for $t \geq 2.5$ sec; as before, $\mu = 155$ Mbps and $\nu \approx \mu/4$.

The MS algorithm is considered in Figure 5.2. The parameters conform to the design rule for pointwise fairing in Section 4.3:

$$\gamma_j = \gamma; \qquad a_j = \sigma_j a, \quad \sigma_1 = \sigma_2 = 1 \quad (j = 1, 2).$$

In Figure 5.2,

$$\begin{aligned}\sum_{j=1}^{2}(a_j/\gamma_j+\nu_j) &= 142 \text{ Mbps for } t<2.5 \text{ sec}\\ &= 284 \text{ Mbps for } t\geq 2.5 \text{ sec.}\end{aligned}$$

Since $\mu = 155$ Mbps it follows that the network is Unsaturated for $t < 2.5$ sec, and Saturated for $t \geq 2.5$ sec. Observe in the figure that for $t < 2.5$ sec, the approach to steady state is non-oscillatory and the queue is essentially empty. After $t = 2.5$ sec, a spike in the queue occurs as the second VC turns on and the first one has not yet given up sufficient bandwidth. Also, for $t > 2.5$ sec, the queue is no longer empty and oscillations in the rates and in the queues exist, but are not excessive. The results in Figure 5.2 indicate that of the two sets of parameters used, the one in (a), which is less cautious, is also the more desirable since the initial ramp-up to steady state is faster, equalization occurs earlier and the incremental cost in increased oscillations is modest.

Now consider the behavior of the JRJ algorithm shown in Figure 5.3. We see similar behavior in the rates and queues. By varying the gain and damping parameters, the transient behavior, amplitude of oscillations and time to convergence in the rates, as well as queue size, can be controlled. Two extreme examples are shown. In (b) the first VC has not yet reached it steady state limit before the second turns on. A major difference between the two algorithms is illustrated in Figures 5.2 and 5.3. As stated earlier, the JRJ algorithm always operates in the Saturated regime. This fact explains the large initial oscillations in rate and spike in the queue in Figure 5.3(a), which was designed to roughly correspond to Figure 5.2(a) in steady state. In order to make the two algorithms operate more similarly, in the remainder of the simulations we have imposed a ceiling $R_{\mathrm{JRJ}}^{\mathrm{ceil}}$ on the rates for the JRJ algorithm, so that for each VC

$$R_{\mathrm{JRJ}}^{\mathrm{ceil}} \simeq \frac{a_{\mathrm{MS}}}{\gamma_{\mathrm{MS}}} + \nu.$$

We have conducted simulations to assess the sensitivity of the algorithms with respect to the queue threshold Q_T and the updating interval Δ. In contrast to Figure 5.2, where the threshold $Q_T = 6.5$ cells, we simulated a system with identical parameters, except that $Q_T = 60.5$ cells. The queue in the latter case is larger, which increases the system response time, but otherwise the behaviors are quite similar.

Similarly we have simulated systems in which the updating interval is four times larger than in Figure 5.2. Although the rates exhibit greater variability and the queues are larger, qualitative behaviors are similar.

5.2 Fairness Issues

In this subsection we provide experimental evidence supporting our analysis of fairness of Sections 4.1–4.5. Let us first consider cases where the design rule of (4.9) is applied with $\sigma_j \equiv 1$, $\nu_j \equiv \nu$ and $J = 2$. The pointwise fairing to be expected in this case is exhibited in Figure 5.2 and also in Figure 5.4.

In Figure 5.2 all the parameters are the same for the two VCs. In particular the feedback round-trip delay τ_j is the same, with $\tau_j = 10$ msec. The rates allocated to the two VCs tend to become identical and the curves overlap perfectly. The same parameters are used in Figure 5.4, except for the round-trip delays, which are now $\tau_1 = 10$ msec and $\tau_2 = 100$ msec; the MS algorithm is used. The responses exhibit pointwise fairing; the rates tend to overlap perfectly when shifted by $\tau_1/2$ and $\tau_2/2$, respectively (recall (4.6)).

Let us next consider evidence of averaged fairness. In Figure 5.5 the MS algorithm is used, and the less restrictive design rule of (4.27) is adopted, with $\sigma_j = 1$ for $j = 1, 2$, and $a_1/\gamma_1 = a_2/\gamma_2 = 25$ Mbps. From inspection of the rates and their difference we note that although no pointwise fairing occurs, the curves oscillate around a common average value, but with different amplitudes. We also obtain similar behavior from the JRJ algorithm.

We have also simulated cases where the round-trip delays are modeled as the sum of a constant component and an exponentially distributed stochastic component. Periodic behavior can not be identified in the responses, but the rates tend to be fair.

We return to the case of constant round-trip delays, and investigate the characteristics of services defined by the choice of parameters ν_j and σ_j (see (4.2)). Figures 5.6(a) and 5.6(b) show implementations of "Best Effort with Minimum Bandwidth" by the MS and JRJ algorithms for $J = 4$, different minimum bandwidths and equal access to uncommitted bandwidth. Figure 5.7 illustrates behavior for another choice of parameters ν_j and σ_j, which is consistant with a more sophisticated service. As the reader may verify, the observed dynamical behaviors converge to conform to the service specifications.

5.3 Adaptability and Interoperability

In order to assess the robustness of our control algorithms with respect to the number of VCs, J, competing for bandwidth at a single node, we observed the impact of increasing J

$\Sigma_J = J$	Mean Queue Length	Std Deviation of Queue Length	Utilization of Network Bandwidth (%)
1	0.472	0.499	47
2	0.943	0.230	94.3
4	6.75	11.37	96
20	79.7	112.00	85
40	181.01	231.16	68.3
100	1262.1	1161	72.1

Table 5.1: Steady state performance with increasing number of VCs J. The MS algorithm was used with $\mu = 45$ Mbps, $\tau = 10$ msec, $\Delta = 5$ msec, $a \cdot \Delta = .21$ Mbps, $\gamma \cdot \Delta = .01$, $Q_T = 4.5$ cells, and $\nu = 0$.

on both their transient and steady state behavior. Recall (4.16), which shows that for the design in (4.9), the behavior of the aggregate bandwidth depends on J via the gain factor $\Sigma_J = \sum_j \sigma_j$.

Figure 5.8 presents the transient behavior of the MS algorithm as new groups of six VCs turn on at 0, 2, 4 and 6 sec, respectively. In the interval up to 2 sec the active VCs do not reach the Saturated regime; after 2 sec the system is Saturated, and the degree of saturation increases at 4 and 6 sec. The rates for a subset of the VCs are shown in Figure 5.8(a), while the aggregate rate and the queue length are shown in Figure 5.8(b). Note the relatively slow growth in the fluctuations as J increases. A selection of results of steady state behavior of the aggregate bandwidth is presented in Table 5.1. The results are obtained by studying the behavior of a single connection under the MS algorithm, when the parameters γ and a are fixed, while the factor Σ_J, chosen for simplicity equal to J, is in the range 1–100. Note that the saturated regime is reached when $\Sigma_J = J > 2$. Larger queues and a degradation of efficiency is noticeable when $J \geq 20$. Rescaling of the parameters of the algorithm may be necessary for larger values of J, as discussed in Section 4.6.

As discussed in Section 4.7, a limited degree of interoperability can be achieved among different algorithms in the class considered in this paper. We have simulated systems with four VCs in which two used the MS algorithm and the remaining two used the JRJ algorithm, and where the service is Best Effort with Minimum Bandwidth. The parameters of the algorithms are selected following (4.49). The responses indeed exhibit fair behavior commensurate with the service definition.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the benefit of several years of collaboration on the topic of this paper with their colleagues Anwar Elwalid, Kerry Fendick, Dick Peck and Vikram Saksena. Special thanks are due to Alan Weiss, who we consulted frequently and who always gave generously of his time and advice.

References

[BME92] L. Benmohamed and S. M. Meerkov, "Feedback control of congestion in packet switching networks: the case of a single congested node," *IEEE/ACM Trans. Networking*, **1**(6), 1993, pp. 693–708.

[BSH90] J.-C. Bolot and A. U. Shankar, "Dynamic behavior of rate-based flow control mechanisms," *ACM SIGCOMM, Comp. Comm. Rev.*, **30**, 1990, pp. 35-49.

[BFE95] F. Bonomi and K. Fendick, "The rate-based flow control framework for the Available Bit Rate ATM service," to appear in the *IEEE Network*, 1995.

[COP65] W. A. Coppel, *Stability and Asymptotic Behavior of Differential Equations*, D. C. Heath, Boston, 1965.

[ELW93] A. I. Elwalid, "Adaptive rate-based congestion control for high speed wide-area networks: stability and optimal design," preprint, 1993.

[FRW92] K. W. Fendick, M. A. Rodrigues and A. Weiss, "Analysis of rate-based feedback control strategy for long haul data transport," *Performance Evaluation*, **16**, 1992, pp. 67–84.

[HYI94] M. Hluchyj and N. Yin, "On closed-loop rate control for ATM networks," *Proc. IEEE INFOCOM 94*, pp. 99–108.

[JAC88] V. Jacobson, "Congestion avoidance and control," *Proc. ACM SIGCOMM'88*, pp. 314–329.

[JAI90] R. Jain, "Myths about congestion management in high-speed networks," Report DEC-TR-726, Digital Equipment Corp., 1990.

[KMR93] H. Kanakia, P. Mishra and A. Reibman, "An adaptive congestion control scheme for real-time packet video transport," *Proc. ACM SIGCOMM*, 1993.

[KES91] S. Keshav, "A control theoretic approach to flow control," *Proc. ACM SIGCOMM*, 1991.

[KES94] S. Keshav, "Packet-pair flow control," preprint, 1994.

[MAC91] B. A. Mackrucki, "On the performance of submitting excess traffic to ATM networks," *Proc. IEEE GLOBECOM '91*, 1991, pp. 281–288.

[MKA92] P. Mishra and H. Kanakia, "A hop by hop rate based congestion control scheme," *Proc. ACM SIGCOMM*, 1992.

[MIT92] D. Mitra, "Asymptotically optimal design of congestion control for high speed data networks," *IEEE Trans. Comm.*, **40** (2), 1992, pp. 301–311.

[MSE90] D. Mitra and J. B. Seery, "Dynamic adaptive windows for high speed data networks: theory and simulations," *Proc. ACM SIGCOMM*, 1990, pp. 20–30.

[MSE93(i)] D. Mitra and J. B. Seery, "Dynamic adaptive windows for high speed data networks with multiple paths and propagation delays," *Computer Networks and ISDN Systems*, **25**, 1993, pp. 663–679. A short version is in Proc. INFOCOM '91, pp. 39–48.

[MSE93(ii)] D. Mitra and J. B. Seery, "New adaptive algorithms for windows and rates in high speed, wide area networks based on periodic, 1-bit, explicit feedback," unpublished report, January 1993.

[MST91] A. Mukherjee and J. C. Strikwerda, "Analysis of dynamic congestion control protocols – a Fokker-Planck approximation," *Proc. ACM SIGCOMM*, 1991, pp. 159–169.

[RJA88] K. K. Ramakrishnan and R. Jain, "A binary feedback scheme for congestion avoidance in computer networks with a connectionless network layer," *Proc. ACM SIGCOMM '88*, pp. 303–313.

[RMI88] K. G. Ramakrishnan and D. Mitra, "PANACEA: An integrated set of tools for performance analysis," in *Modelling Techniques and Tools for Computer Evaluations*, Ed. R. Puigjaner and D. Potier, Plenum Press, NY, 1988, pp. 25–40.

[SEE88] J. B. Seery, "Customizing PANACEA/Simulation," *AT&T Bell Laboratories Memorandum*, 1988.

[ZHA89] L. Zhang, "A new architecture for packet switching network protocols," Ph.D. Thesis, MIT, Lab. for Comp. Sci., Cambridge, MA 02139, 1989.

[ZHA90] L. Zhang, "Virtual clock: a new traffic control algorithm for packet switching network," *Proc. ACM SIGCOMM*, 1990, pp. 8–18.

Figure 5.1:

Figure 5.2:

Figure 5.3:

Figure 5.4:

Figure 5.5:

Figure 5.6:

Figure 5.7:

Figure 5.8:

Figure 5.9:

Figure 5.10:

Figure 5.11:

Figure 5.12:

Figure 5.13: