PAPER

Stability Analysis for Global Performance of Flow Control in High-Speed Networks Based on Statistical Physics

Masaki AIDA[†] *and* **Kenji HORIKAWA**[††], *Members*

SUMMARY This paper focuses on flow control in high-speed and large-scale networks. Each node in the network handles its local traffic flow only on the basis of the information it knows. It is preferable, however, that the decision making of each node leads to high performance of the whole network. To this end, the relationship between local decision making and global performance of flow control is the essential object. We propose phenomenological models of flow control of high-speed and large-scale networks, and investigate the stability of these models.

key words: *high speed network, flow control, autonomous distributed system, throughput, Fokker-Planck equation*

1. Introduction

This paper investigates performance and stability of flow control schemes in high-speed and large-scale networks. In order to clarify our motivation, this section states the characteristics of high-speed and large-scale networks and states issues concerning the frameworks of network control associated with them.

1.1 Issues in High-Speed Networks

In a high-speed network, propagation delay becomes a dominant factor in the transmission delay. This is because light speed is a non-scaling factor, and is the same for high-speed networks. Therefore, at a given time, a large amount of data is being propagated on links in the network (Fig. 1). The amount of this data is characterized by *delay-bandwidth product*, that is, the propagation distance times transmission rate. Therefore, in high-speed and/or long-distance transmission, there is a larger amount of data on links than in nodes. Figure 2 shows an example of how much data can be on a link. Let us consider the situation involving data transmission between two nodes, with a distance between them of 1 km and a link speed of 1 Mbps. If transmission speed increases to 1 Gbps, the data amount on the link is equivalent to 10^3 km on a 1-Mbps link. In addition, if its transmission speed increases to 1 Tbps, the data amount is equivalent to 10^6 km on a 1-Mbps link. This distance is about 2.5 times the distance between the earth and the moon.

This means that it is impossible to exert time-sensitive control based on collecting global information about the network. So, the frameworks of time-sensitive control, in a large-scale and high-speed network, are inevitably autonomous distributed systems.

1.2 Issues in Large-Scale Networks

Let us consider an approach analyzing a large-scale network. If we model it based on traditional queueing network theory, a lot of degrees of freedom of the network system are included in the model. For example, in order to derive the global performance from the state of each node, it is necessary to obtain the joint probability distribution of the states of all nodes in the network. If the distribution is easily determined, there is no problem. However, it is actually impossible to obtain the distribution of a lot of nodes including their correlated

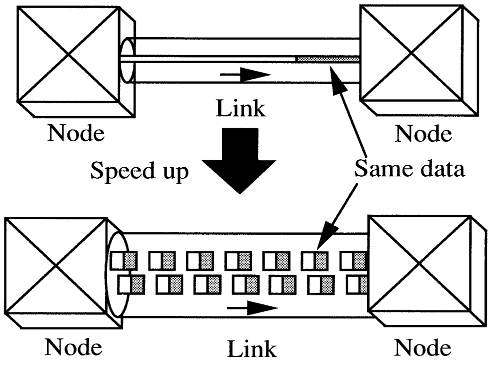


Fig. 1 Large delay-bandwidth product.

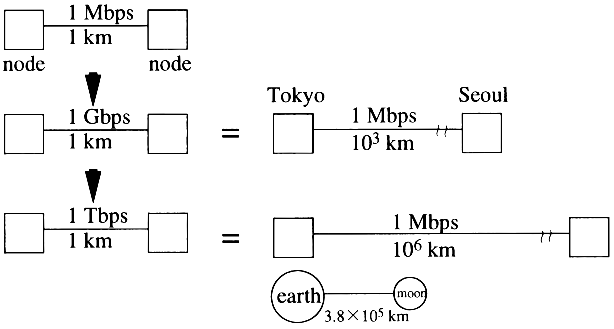


Fig. 2 An example of large delay-bandwidth product.

Manuscript received December 11, 1998.
Manuscript revised June 24, 1999.

[†] The author is with Traffic Research Center, NTT Advanced Technology Corporation (NTT-AT), Musashino-shi, 180-0006 Japan.

[††] The author is with NTT Information Sharing Platform Laboratories, Musashino-shi, 180-8585 Japan.

effects.

When all nodes are mutually independent, as in cases of BCMP theorem [2], the probability distribution is obtained in a simple product-form. These cases, however, are strongly restricted situations. Conversely, if we assume the independence among nodes, the interesting behavior of global performance can not be described. Thus, the approach directly describing all degrees of freedom is difficult to handle.

In general, when a system is in equilibrium, the system is characterized by a few order parameters. The kind of these order parameters is much smaller than the degrees of freedom of the whole system. Therefore we adopt an approach which directly describes the behavior of a global performance as an order parameter. Then the effect of the behavior of each node is reflected as *fluctuation* of the global performance.

1.3 Framework of Flow Control in High-Speed and Large-Scale Networks

This paper focuses on flow control in networks in which nodes handle their local traffic flow themselves based only on the information they know. It is, of course, preferable that the decision making of each node leads to high performance of the whole network. In flow control, we adopt the whole throughput of a network as a global performance measure.

To achieve such *coherent* flow control, the relationship between the local decision and the global performance of flow control is the essential object. We propose phenomenological models of flow control for a large-scale and high-speed network, and investigate the stability of these models. These models assume that the whole throughput of a network has a Markovian property, and they use the technique of Ω-expansion of the master equation. Based on these models, we have tried to connect the whole throughput and its fluctuation caused by the local decision making of each node.

2. Background

2.1 Methods in Statistical Physics and Thermodynamics

Data in a network is in one of the following two states: in a node, or on a link. We define the throughput of a network at time t as how much data is being propagated on the network [3], [6]. So, the throughput of a network is the amount of data on the links at time t, and is denoted by $X_{\mathrm{E}}(t)$. Alternately, the average link utilization of the whole network, that is, the normalized value of the amount of data on the links, may also be thought of as the throughput of the network. We denote it by $X_{\mathrm{I}}(t)$.

These two quantities are categorized from the thermodynamical point of view as follows:

- Extensive Quantity
  Consider an equilibrium system made by combining n identical equilibrium subsystems. When the thermodynamical quantity of the large system equals n times the quantity of one small subsystem, the quantity is called an extensive quantity. $X_{\mathrm{E}}(t)$ is a quantity of this type.
- Intensive Quantity
  In the same situation, when the thermodynamical quantity of the large system equals the quantity of one small subsystem, the quantity is called an intensive quantity. $X_{\mathrm{I}}(t)$ is a quantity of this type.

These quantities are homogeneous functions of degree Ω^1 and Ω^0, respectively, where Ω is system size and, in this case, denotes the maximum amount of data on links in the whole network.

The behavior of throughput X_{E} is determined by the behaviors of all nodes in the network. However, when the network size Ω is large, there are many nodes in the network and it is almost impossible to describe their behavior, including the interactions among all nodes. Therefore, we assume that the influence of the behavior of a node is reflected in a small fluctuation of X_{E}. So we regard X_{E} as a random variable.

We define the probability density function $P_{\mathrm{E}}(x,t)$ such that the probability of the throughput $X_{\mathrm{E}}(t)$ being in $x \leq X_{\mathrm{E}}(t) < x + dx$ at time t is $P_{\mathrm{E}}(x,t)\,dx$. We assume $X_{\mathrm{E}}(t)$ as a Markovian random variable, and describe the temporal evolution of $P_{\mathrm{E}}(x,t)$ using the master equation:

$$\frac{\partial}{\partial t} P_{\mathrm{E}}(x,t) = \int W(x-r,r,t)\, P_{\mathrm{E}}(x-r,t)\, dr - \int W(x,r,t)\, P_{\mathrm{E}}(x,t)\, dr, \tag{1}$$

where $W(x,r,t)$ denotes the transition rate from $X_{\mathrm{E}}(t) = x$ to $x + r$ at time t. Our interest is in the situation for large networks, so we assume the following conditions:

$$\Omega \gg 1, \;\; E(X_{\mathrm{E}}) \gg 1. \tag{2}$$

2.2 Ω-Expansion of the Master Equation

Let us define moments of $W(x,r,t)$ with respect to transition r as

$$C_n(x,t) := \int r^n\, W(x,r,t)\, dr. \tag{3}$$

If all moments $C_n(x,t)$ exist for all n, master equation (1) is written as

$$\frac{\partial}{\partial t} P_{\mathrm{E}}(x,t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \left(\frac{\partial}{\partial x}\right)^n C_n(x,t)\, P_{\mathrm{E}}(x,t). \tag{4}$$

This is called Kramers-Moyal expansion [4], [5].

In order to describe the relationship between the temporal evolution of $X_{\rm E}$ and its fluctuation, we use Ω-expansion (size expansion).

When the network size Ω is very large and $X_{\rm E}$ is near equilibrium, we assume

$$E(X_{\rm E}) = \mathcal{O}(\Omega),\ \ Var[X_{\rm E}] = \mathcal{O}(\Omega), \tag{5}$$

where $Var[\cdot]$ denotes the variance. Here, we introduce normalized throughput $X_{\rm I}$ (intensive quantity) as

$$X_{\rm I}(t) := \frac{X_{\rm E}(t)}{\Omega}, \tag{6}$$

which is independent of the network size. The variance of $X_{\rm I}$ is then expressed as

$$Var[X_{\rm I}] = \varepsilon^2\, Var[X_{\rm E}], \tag{7}$$

where $\varepsilon := \Omega^{-1}$. We define the probability that $X_{\rm I}(t)$ is in $x \le X_{\rm I}(t) < x + dx$ at time t as $P_{\rm I}(x,t)\,dx$. From the normalization condition of $P_{\rm I}(x,t)$, we obtain

$$P_{\rm I}(x,t) = \Omega\, P_{\rm E}(\Omega x, t), \tag{8}$$

under scaling law (6). The transition rate, $w(x,r,t)$, from $X_{\rm I} = x$ to $x + r$ at time t is

$$w(x,r,t) = \varepsilon\, W(\Omega x, r, t). \tag{9}$$

The physical meaning of this scaling is that the transition r in extensive quantity $X_{\rm E}$ has Ω times the opportunity to occur as the same transition in normalized intensive quantity $X_{\rm I}$. Using these scaling rules, we rewrite master equation (1) in the following Kramers-Moyal expansion form:

$$\begin{aligned}\frac{\partial}{\partial t}\, P_{\rm I}(x,t) \;&=\; \sum_{n=1}^{\infty} \frac{(-1)^n}{n!}\, \varepsilon^{n-1} \\ &\quad \times \left(\frac{\partial}{\partial x}\right)^n c_n(x,t)\, P_{\rm I}(x,t), \end{aligned}\tag{10}$$

where $c_n(x,t)$ denotes the n-th moment of $w(x,r,t)$ with respect to transition r, as

$$c_n(x,t) = \int r^n\, w(x,r,t)\, dr. \tag{11}$$

Equation (10) is called the Ω-expansion [4], [5] of master equation (1).

2.3 Cumulant Expansion and Temporal Evolution of Cumulants

We define the characteristic function of $P_{\rm I}(x,t)$ as

$$Q(\xi,t) := \int P_{\rm I}(x,t)\, e^{i\xi x}\, dx, \tag{12}$$

and the characteristic function of the transition rate as

$$\omega(\xi,r,t) := \int w(x,r,t)\, e^{i\xi x}\, dx. \tag{13}$$

The master equation of $Q(\xi,t)$ is then expressed as

$$\begin{aligned}\frac{\partial}{\partial t}\, Q(\xi,t) \;&=\; \frac{1}{2\pi} \int dr \int d\eta \int_0^r ds\, i\xi\, e^{i\varepsilon\xi s} \\ &\quad \times Q(\xi-\eta,t)\, \omega(\eta,r,t). \end{aligned}\tag{14}$$

We assume that the solution $Q(\xi,t)$ of master equation (14) has the following form:

$$Q(\xi,t) = \exp\, q(\xi,t) = \exp\left[\sum_{n=1}^{\infty} \frac{(i\xi)^n}{n!} q_n(t)\right]. \tag{15}$$

Here, $q_n(t)$ is the n-th cumulant of $X_{\rm I}$.

From Eq. (14), we can derive the temporal evolution of cumulants $q_n(t)$s. $q_1(t)$ and $q_2(t)$ respectively correspond to the average and the variance of $P_{\rm I}(x,t)$. Expanding $q_1(t)$ up to $\mathcal{O}(\varepsilon)$ with respect to the power of ε, we define $y(t)$, $u(t)$ as

$$q_1(t) = y(t) + \varepsilon\, u(t) + \mathcal{O}(\varepsilon^2). \tag{16}$$

Similarly, expanding $q_2(t)$ up to $\mathcal{O}(\varepsilon)$ with respect to the power of ε, we define $v(t)$ as

$$q_2(t) = \varepsilon\, v(t) + \mathcal{O}(\varepsilon^2). \tag{17}$$

The temporal evolution equations of $y(t)$, $v(t)$, $u(t)$ are, as shown in [5], expressed as

$$\frac{\partial y(t)}{\partial t} \;=\; c_1(y,t), \tag{18}$$

$$\frac{\partial v(t)}{\partial t} \;=\; 2\, c_1'(y,t)\, v(t) + c_2(y,t), \tag{19}$$

$$\frac{\partial u(t)}{\partial t} \;=\; c_1'(y,t)\, u(t) + \frac{1}{2}\, c_1''(y,t)\, v(t), \tag{20}$$

where c_n' and c_n'' denote the first and second derivatives, respectively, of c_n with respect to y. Note that the temporal evolution of the cumulants is determined by the moments of the transition rate.

3. Flow Control Schemes and Simulation Results

3.1 Network and Node Models

Our network model has simple lattice topology and a torus boundary, that is, a closed Manhattan Street network (Fig. 3(a)). There are 400 nodes in the network and they have a 20×20 lattice configuration. All nodes have two incoming links and two outgoing ones. Two links of a node correspond to the vertical and horizontal directions. For simplicity, we assume an Asynchronous Transfer Mode (ATM) network in which the data unit, the cell, has a fixed length in bits.

All nodes have a switching function such that the incoming cells are switched into the vertical or horizontal outgoing direction with the probability 1/2 based on a Bernoulli trial. Switched cells are stored in the

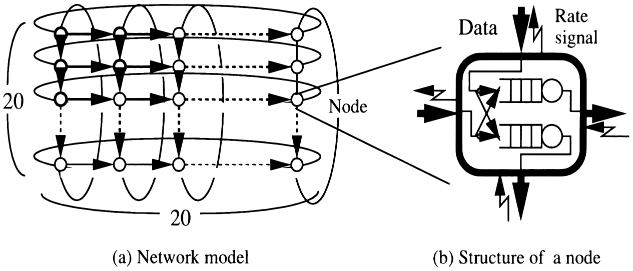


Fig. 3 Network and node models.

output buffer of their direction (Fig. 3(b)).

All nodes are capable of receiving and sending rate signals. When a node receives a rate signal from downstream, the node adjusts its transmission rate to downstream according to the rate signal. In addition, all nodes can signal the allowed transmission rate to the upstream nodes by using a rate signal. Each node autonomously decides how much rate should be specified by using a rate signal. Different rate control schemes are applied in accordance with the different processes used to determine the specified rate.

Here, we state supplements of characteristics in the network model.

- **Network topology has high symmetry**
  Although the homogeneous topology of the network model may look un-realistic, the reason is as follows. If we can identify the bottleneck points, in performance, in the network model, then we can focus on the part of the network model including the bottleneck points. However, if the symmetry of the network topology is high and the bottleneck points in performance are not explicitly identified, the above model focusing on part of the network can not be applied. Actually, any parts can be bottlenecks with the same probability. Let us consider the situation when accidental congestion occurs at a certain node. Then we are interested in the behavior of the local congestion, whether

  – it grows and causes deterioration of the whole network performance through interaction among nodes, or
  – it remains a local phenomenon and diminishes with time.

  In order to evaluate the behavior without other complicated effects, we adopt the above simple network model. This is the reason why we do not take hierarchical network models (or, for example, additional flow control between terminals such as TCP) into consideration.

- **All nodes obey the same rule**
  There are no special nodes such as one to control the whole network. More specifically, such nodes can not exist under the high-speed network environment. We therefore give all nodes the same rule. Note that although the rule is the same, each node is in a different state at some time t and thus the behavior of each node at t is different. We are interested in the behavior of global performance under the situation that

  – all nodes behave autonomously by the same rule, and
  – there is no special node controlling the whole network.

- **Network size and boundary condition**
  There is no essential reason that the number of nodes is 400 with a torus boundary. It is only necessary that there be as many nodes as possible, and the nodes on the network boundary behave under the same rule and conditions as the other ordinary nodes. The reason for a closed network is that the number of cells in the network should be invariant. This is required, to compare the different flow control schemes under the same conditions, as shown in below. Note that if we focus on a part of the network, it can be regarded as an open network in which the number of cells is not invariant.

3.2 Two Flow Control Schemes

In this subsection, we show two different rate control schemes for the different processes used to decide the specified rate.

One is for specifying rates using a rate signal from downstream, i.e., a rate-based flow control. We call it Rate Driven Control (RDC). Ordinarily, each node sends a rate signal at every deterministic time interval. Let D be the round-trip time between adjacent nodes. For simplicity, we set the time interval between consecutive rate signals to D.

Consider a node downstream from link i. The node calculates the following quantity, at every time period D:

$$\widetilde{N}_i(t) = N_i(t) + \{R_r^{\rm in}(i, t-D) - R_r^{\rm out}(i, t-D)\}\, D, \tag{21}$$

where $R_r^{\rm in}(i,t)$ denotes the specified rate to send upstream at time t, $R_r^{\rm out}(i,t)$ denotes the specified rate received from downstream at time t, and $N_i(t)$ denotes the queue length of the buffer. Then the node calculates the new specified rate as

$$R_r^{\rm in}(i,t) = R_r^{\rm in}(i, t-D) + \frac{\alpha L - \widetilde{N}_i(t)}{D}, \tag{22}$$

where L denotes the buffer capacity, and α is a threshold parameter such that αL denotes the threshold of the buffer. Ordinarily, each node sends a rate signal at every deterministic time interval, D. However, in case of emergency, that is, during buffer overflow/underflow,

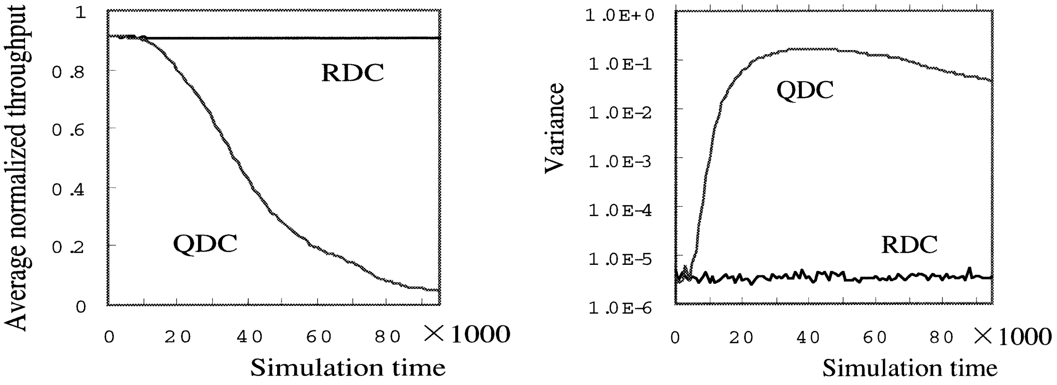


Fig. 4 Behavior of normalized throughput.

each node can immediately send a rate signal.

The other scheme specifies a rate based only on the queue length of the nodes' own buffers, i.e., a queue-threshold based flow control. We call it Queue Driven Control (QDC). Ordinarily, each node sends a rate signal at every deterministic time interval, D. The specifying rate is determined by the queue length of its buffers. Each node makes an effort to ensure that the queue length is equal to the predefined threshold in its buffers. Let $R_q^{\text{in}}(i,t)$ be the specified rate to send upstream at time t:

$$R_q^{\text{in}}(i,t) = R_q^{\text{in}}(i,t-D) + \frac{\alpha L - N_i(t)}{D}. \qquad (23)$$

As with RDC, in case of emergency, that is, during buffer overflow/underflow, each node can immediately send a rate signal.

3.3 Experimental Evaluation of Stability

Let us investigate stabilities of the above two flow control schemes, RDC and QDC, by using a Monte Carlo simulation. The environment of our simulation is as follows:

- **bandwidth and distance between adjacent nodes**
  Each horizontal link in Fig. 3(a) is 300 Mbps and 20 km in length, and each vertical link is 600 Mbps and 2 km in length. The links are assumed to be WANs and LANs, respectively. Note that the horizontal links are the bottle-necks of the traffic flow.
- **buffer capacity**
  All nodes have the same capacity, 300 cell places. The threshold parameter α is set to be 1/2.
- **cell overflow**
  In order to maintain the network load, the discarded cells, when no buffer space is left, are again added to network nodes chosen at random. Therefore, the total number of cells in the network is invariant.
- **total number of cells**
  There are 40,000 cells in the network. This corresponds to about 160% of the total capacity of all links. This implies that the load of the network is high.

Based on the above environment, we compare the throughput and queue length of the two networks, one controlled by RDC and the other by QDC.

At the initial time $t = 0$, both networks are near equilibrium states. Figure 4 shows the behavior of the average and the variance of the normalized throughput of RDC and QDC. These figures show that RDC is stable but QDC is not, and the variance is remarkably increased for QDC.

Figure 5 shows the queue lengths of buffers in the nodes in the network. These data are from a typical sample path of the simulation. Each pixel corresponds to a node and the configuration is the same as a network topology. The queue length of a node is denoted by the color density. "Step" denotes simulation time, "Link" denotes the normalized throughput, "Queue" denotes the average queue length, and "Loss" denotes the number of lost cells (running total).

The figure shows that the queue length remains homogeneously distributed for RDC. However, for QDC, node queue lengths go to basically two states: very long and almost empty. The nodes with very long queue lengths are clustered. This clustering process corresponds to the process of decreasing normalized throughput, shown in Fig. 4.

3.4 Experimental Evaluation of Robustness

Next, let us investigate robustness of the performance with respect to increasing network load. Simulation environments are almost the same as the above model, but the number of cells in the networks are increased with time. The number of cells in the network is zero at the initial state. Then, five cells are added to network nodes chosen at random, every one simulation time.

The results obtained in the simulation are shown in Fig. 6. Note that the number of cells increases as

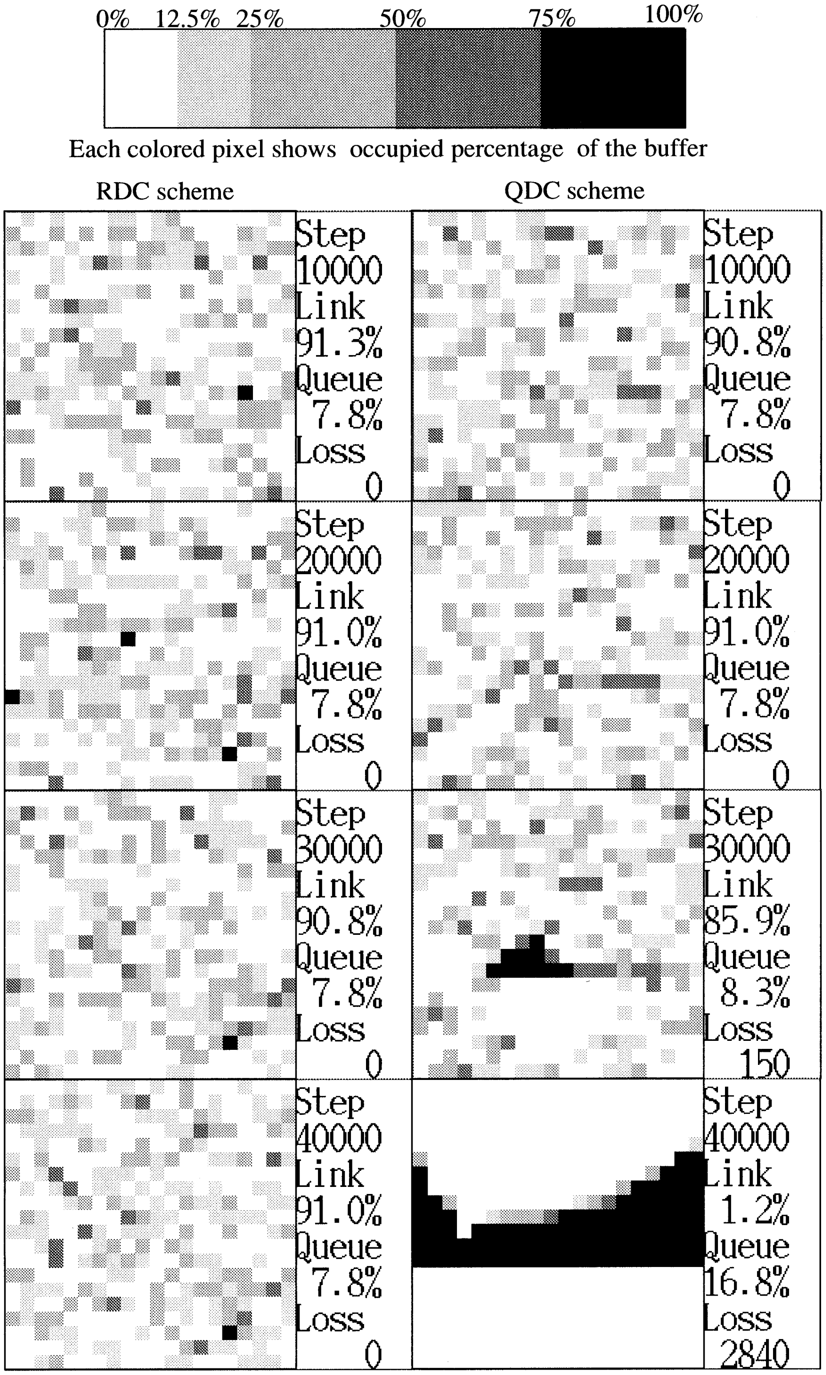
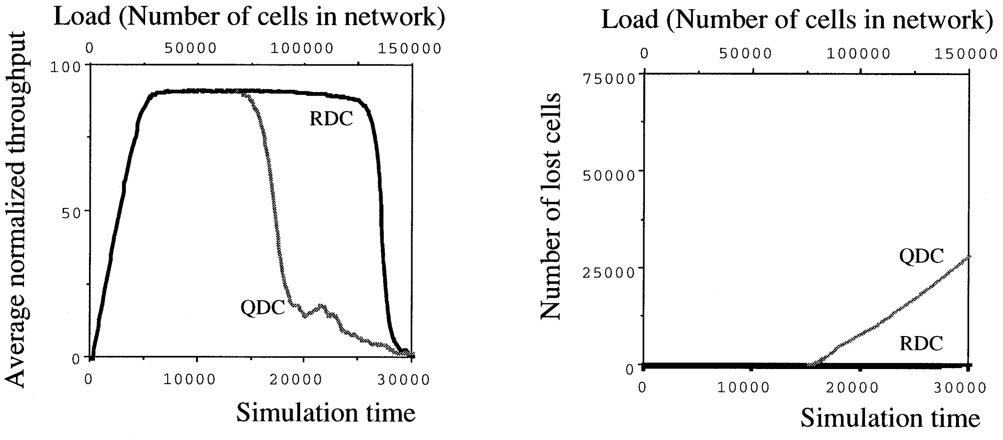

0%
12.5%
25%
50%
75%
100%
Each colored pixel shows occupied percentage of the buffer
RDC scheme
QDC scheme
Step 10000 Link 91.3% Queue 7.8% Loss 0
Step 10000 Link 90.8% Queue 7.8% Loss 0
Step 20000 Link 91.0% Queue 7.8% Loss 0
Step 20000 Link 91.0% Queue 7.8% Loss 0
Step 30000 Link 90.8% Queue 7.8% Loss 0
Step 30000 Link 85.9% Queue 8.3% Loss 150
Step 40000 Link 91.0% Queue 7.8% Loss 0
Step 40000 Link 1.2% Queue 16.8% Loss 2840

Fig. 5 Behaviors of queue length in each node.

time progresses. As shown in Fig. 6(a), most of cells remain on the link between 0 and 4,000 cell times because the network is not busy, and incoming cells are immediately sent out. After 5,000 cell times, the throughput becomes constant at about 92%. This means that cells begin to be stored in buffers.

After 15,000 cell times (after the number of cells in the network exceeds 75,000), the throughput of the network controlled by the QDC begins to drop. Finally, its throughput is lowered to nearly 0 and the control fails. On the other hand, even at 25,000 cell times (number of cells is 125,000), RDC maintains high throughput without failure. In addition, RDC causes hardly any cell loss, as shown in Fig. 6(b). The throughput RDC begins to drop and the control is fails at 30,000 (150,000 cells). Since the total capacity of the buffer for the bottle neck links is 120,000 cells, however, it is safe to say that RDC can endure in higher load conditions and can maintain normal network operation.

Figure 7 shows the behaviors of the queue lengths



(a) Throughput of network

(b) Number of lost cells

Fig. 6 Behaviors of throughput and the number of lost cells.

of buffers in the nodes in the networks as the same in Fig. 5. Since the number of cells in the network increases as the simulation advances, it becomes blackish overall. It is possible to see that the areas of congested nodes spread in the QDC. On the other hand, in the RDC it can be understood that there are some congested nodes but they are not clustered, and the expansion of congestion is minimized to prolong control without cell loss.

4. Linear Relaxation Model of Rate

4.1 Linear Interaction Model

This subsection proposes a phenomenological model for RDC. Let $M(i,t)$ be the throughput of link i, that is, the number of cells on link i at time t. We assume the utilization of the network is sufficiently high.

When $R_r^{\rm in}(i,t) = R_r^{\rm in}(i,t-D)$, no change occurs. If $R_r^{\rm in}(i,t) \neq R_r^{\rm in}(i,t-D)$, $M(i,t)$ changes to $\frac{1}{2}DR_r^{\rm in}(i,t)$. The time for achieving $\frac{1}{2}DR_r^{\rm in}(i,t)$ is independent of its value, and instead depends only on the round-trip time D, that is,

$$\begin{aligned} &M(i,t+D) - M(i,t) \\ &\quad = \frac{1}{2}\,D\,(R_r^{\rm in}(i,t) - R_r^{\rm in}(i,t-D)). \end{aligned} \tag{24}$$

This means the difference of $R_r^{\rm in}(i,t)$ and $R_r^{\rm in}(i,t-D)$ determines the transition rate of $M(i,t)$. When $M(i,t)$ is near equilibrium $M_0(i,t)$, and since we regard D as a time constant, we model

$$\frac{\partial}{\partial t}\,M(i,t) = -\frac{M(i,t) - M_0(i,t)}{D}, \tag{25}$$

as linear relaxation.

The throughput of the network is then denoted by

$$X_{\rm E}(t) = \sum_i M(i,t). \tag{26}$$

4.2 Gaussian Approximation near Stable Equilibrium State

Let the normalized throughput be $X_{\rm I} = y_0$ when the network is in equilibrium. From Eqs. (6), (25) and (26), we have

$$\begin{aligned} &X_{\rm I}(t+\Delta t) - X_{\rm I}(t) \\ &\quad = -\frac{X_{\rm I}(t) - y_0}{D}\,\Delta t + \mathcal{O}(\varepsilon)\,\Delta t, \end{aligned} \tag{27}$$

for small Δt. Since $y(t)$ in Eq. (16) means the leading term ($\mathcal{O}(1)$ term) in the first cumulant (the average) of $X_{\rm I}$, we focus only on $\mathcal{O}(1)$ term in Eq. (27) and replace $X_{\rm I}(t)$ with $y(t)$. Then we have

$$\frac{\partial y(t)}{\partial t} = -\frac{y - y_0}{D}. \tag{28}$$

From the temporal evolution equation (18), we have

$$c_1(y,t) = -\frac{y - y_0}{D}, \tag{29}$$

as shown in [1]. In addition, we assume

$$c_2(y,t) = \text{const.} = b_r, \tag{30}$$

like Brownian motion. We set the initial distribution of $X_{\rm I}$ as

$$P_{\rm I}(x,0) = \delta(x - x_0). \tag{31}$$

Then non-equilibrium distribution of $X_{\rm I}$ near the equilibrium state is

$$\begin{aligned} P_{\rm I}(x,t) \;=\; &\frac{1}{\sqrt{2\pi\varepsilon v(t)}} \\ &\times \exp\left[-\frac{(x - y_0 - (x_0 - y_0)\,e^{-t/D})^2}{2\,\varepsilon\,v(t)}\right], \end{aligned} \tag{32}$$

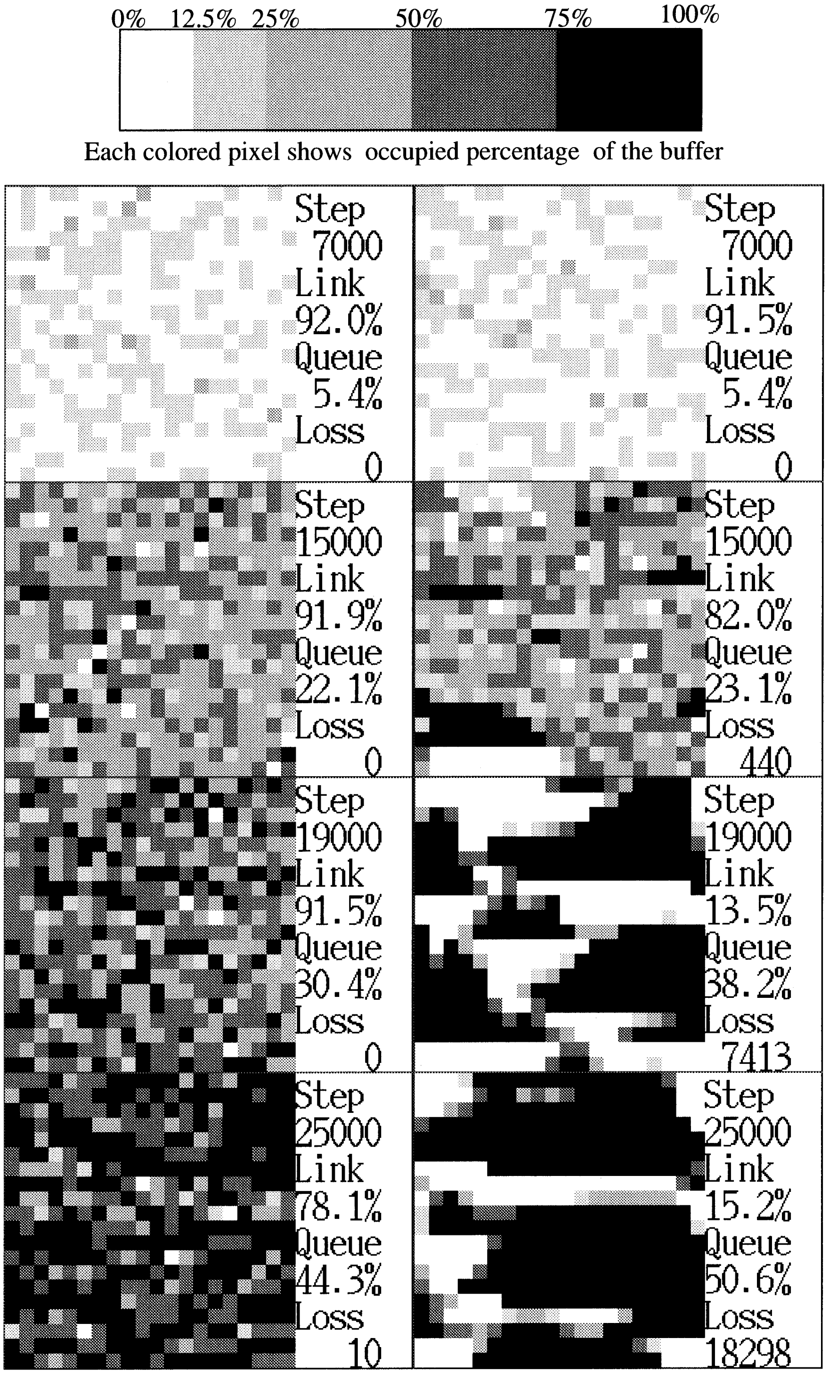


Fig. 7 Behaviors of queue length in each node.

where, from $v(0) = 0$,

$$v(t) = \frac{b_r D}{2}\,(1 - e^{-2t/D}), \tag{33}$$

this is the solution of Eq. (19). The non-equilibrium distribution (32) is the solution of the Fokker-Planck equation [7]

$$\frac{\partial}{\partial t}\,P_{\rm I}(x,t) = \left[-\frac{\partial}{\partial x}\,c_1(x,t) + \varepsilon\,\frac{1}{2}\,\frac{\partial^2}{\partial x^2}\,c_2(x,t)\right] P_{\rm I}(x,t), \tag{34}$$

obtained by truncating Eq. (10) after $\mathcal{O}(\varepsilon^2)$.

Figure 8 shows an example of temporal evolution of the solution (32), where we set $y_0 = 0$, $x_0 = -1$, $D = 1$, and $\varepsilon = b_{\rm r} = 1$ for simplicity. Initially, the solution is around x_0, and afterward it moves to y_0 as t increases. The movement is slowing down when the

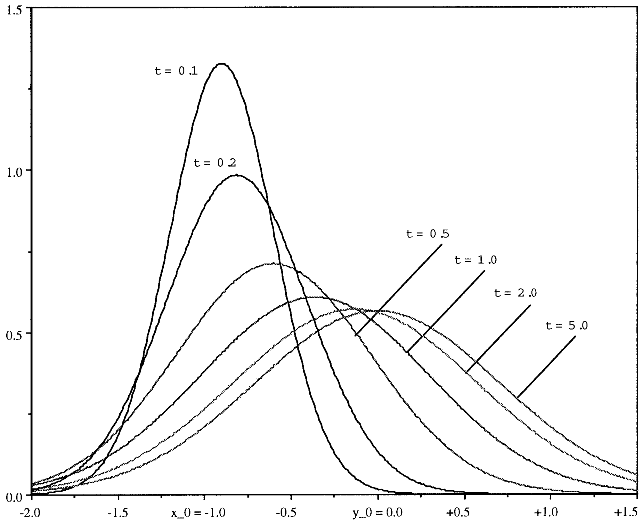


Fig. 8 Behavior of the solution of Gaussian approximation.

average of the distribution (32) approaches the equilibrium y_0. The distribution (32) approaches, in accordance with increasing t,

$$\lim_{t\to\infty} P_{\rm I}(x,t) = \frac{1}{\sqrt{\pi\varepsilon b_r D}} \exp\left[-\frac{(x-y_0)^2}{\varepsilon\, b_r D}\right], \quad (35)$$

and it is independent of the initial state $X_{\rm I} = x_0$.

5. Nonlinear Relaxation Model of Queue Length

5.1 Redefinition of a Network Performance Measure

This subsection shows a model of the QDC scheme. For QDC, the behavior of the throughput is directly controlled by the queue length in each node through (23). Since the time derivative of the throughput corresponds to the second time derivative of the queue length, the equation describing temporal evolution of the throughput is, therefore, the second differential equation with respect to time. This implies that the solution of the equation contains some oscillation modes. It is thought that such oscillation causes the deterioration of throughput shown in Fig. 5. Unfortunately, the technique based on the master equation is not applicable to this type of equation.

To avoid this, we want to focus on the trend of the deterioration process of throughput without oscillation. To this end, we describe the behavior of the queue length in each buffer directly using a phenomenological model. Since, in our network model, there is a constant number of cells in the closed network, if we know the behavior of queue lengths, we can find the throughput, that is, the number of cells on links. For simplicity, we introduce the parameterized buffer shown in Fig. 9. Let us assign 0 to the point of half the buffer capacity, that is, just at the threshold. The fully occupied and emptied points of the buffer are assigned $+\frac{1}{2}L$ and $-\frac{1}{2}L$, respectively. Values of $N_i(t)$ are denoted by this buffer.

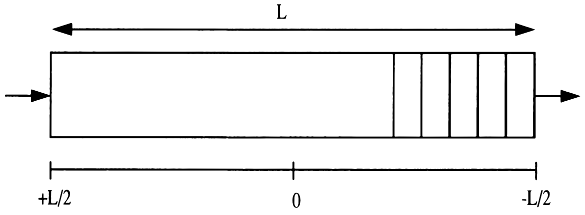


Fig. 9 Buffer model.

We define the total number of cells in all buffers in the network at time t, denoted by $N_{\rm E}(t)$, and the normalized value of $N_{\rm E}(t)$ denoted by $N_{\rm I}(t)$, that is,

$$N_{\rm E}(t) := \sum_i N_i(t), \quad (36)$$

$$N_{\rm I}(t) := N_{\rm E}(t)/\Omega_N, \quad (37)$$

where Ω_N is system size and, in this case, denotes the number of buffers in the network. $N_{\rm E}(t)$ and $N_{\rm I}(t)$ are categorized into extensive and intensive quantity, respectively, same as $X_{\rm E}(t)$ and $X_{\rm I}(t)$.

We define the probability density function $\varphi(\chi,t)$ such that the probability of the normalized queue length $N_{\rm I}(t)$ being in $\chi \le N_{\rm I}(t) < \chi + d\chi$ at time t is $\varphi(\chi,t)\,d\chi$. We assume $N_{\rm I}(t)$ as a Markovian random variable, and describe the temporal evolution of $\varphi(\chi,t)$ using a master equation. Truncating the Ω-expansion of the master equation after $\mathcal{O}(\epsilon^2)$, the temporal evolution equation of $\varphi(\chi,t)$ is obtained as the following Fokker-Planck equation:

$$\frac{\partial}{\partial t}\varphi(\chi,t) = \left[-\frac{\partial}{\partial\chi} c_1(\chi,t) + \epsilon\,\frac{1}{2}\,\frac{\partial^2}{\partial\chi^2} c_2(\chi,t)\right] \varphi(\chi,t), \quad (38)$$

where $\epsilon := \Omega_N^{-1}$, and c_1 and c_2 are drift and diffusion coefficients, respectively.

5.2 Minimal Nonlinear Interaction Model

As shown in Sect. 3.3, the QDC scheme is unstable in equilibrium at high throughput. In addition, complicated behavior such as clustering implies that there is nonlinear interaction among the nodes. It is thus necessary to consider a nonlinear differential equation.

Let us consider the minimal nonlinear interaction describing the QDC scheme. We consider the Fokker-Planck equation (38) and assume the nonlinearity is only in the drift coefficient $c_1(\chi,t)$, and the diffusion coefficient $c_2(\chi,t)$ describes the simple Gaussian white noise.

There are three equilibrium states of $N_{\rm I}(t)$:

$$N_{\rm I}(t) = 0,\ \ N_{\rm I}(t) = +\frac{1}{2}L,\ \ N_{\rm I}(t) = -\frac{1}{2}L. \tag{39}$$

As shown in Sect. 3.3, $N_{\rm I}(t) = 0$ is unstable.

We assume $c_1(\chi, t)$ is to have the following power series expansion form:

$$c_1(\chi, t) = \sum_{n=0}^{\infty} a_n(t)\,\chi^n. \tag{40}$$

Since the buffer threshold, that is, the target queue length, is at half the buffer capacity, $a_n(t) = 0$ for all even n. Thus, the terms of the even degrees of χ are out of consideration from the symmetry of the buffer. In addition, if $X_{\rm I}$ is near equilibrium, i.e., $\chi \simeq 0$, the terms of the higher degrees of χ are negligible. Therefore, the drift coefficient (40) is approximately denoted as

$$c_1(\chi, t) \simeq a_1(t)\,\chi + a_3(t)\,\chi^3, \tag{41}$$

near equilibrium $\chi \simeq 0$. Since $N_{\rm I}(t) = 0$ is unstable, let us choose $a_1(t) > 0$.

If we assume the simplest non-linear interaction among nodes, $c_1(\chi, t)$ is assumed to have the form (41) for all χ. Since $\chi = 0, \pm L/2$ are equilibrium points and $c_1(\chi, t) = 0$ at the equilibrium points, we have

$$c_1(\chi, t) = \gamma\,\chi - g\,\chi^3, \tag{42}$$

where γ, $g > 0$ and

$$g = 4\gamma/L^2. \tag{43}$$

It is natural to consider that γ is determined by the round-trip time D.

On the other hand, since the diffusion coefficient $c_2(\chi, t)$ is assumed to describe the simple Gaussian white noise, we have $c_2(\chi, t) = \text{const.} = b_q$.

The minimal nonlinear Fokker-Planck equation is then obtained by

$$\frac{\partial}{\partial t}\,\varphi(\chi, t) = \left[\frac{\partial}{\partial \chi}\,(-\gamma\,\chi + g\,\chi^3) + \frac{\epsilon b_q}{2}\,\frac{\partial^2}{\partial \chi^2}\right]\,\varphi(\chi, t). \tag{44}$$

In addition, we denote the first and second cumulants of $\varphi(\chi, t)$ as Eqs. (16) and (17), after we replace ε with ϵ. Then, the temporal evolution equations of the cumulants of $\varphi(\chi, t)$ are obtained as Eqs. (18)–(20).

5.3 Nonlinear Scaling and Scaling Solution

Let the initial distribution of the nonlinear Fokker-Planck equation (44) be $\varphi(\chi, 0) = \delta(\chi)$. This is the unstable equilibrium. According to the nonlinear scaling theory [8], the solution of (44) is obtained approximately as

$$\varphi(\chi, t) \simeq \frac{1}{\sqrt{2\pi\tau}}\left[1 - \frac{g}{\gamma}\,\chi^2\,(1 - e^{-2\gamma t})\right]^{-3/2} \times \exp\left\{-\frac{\chi^2}{2\,\tau\,[1 - (g/\gamma)\,\chi^2\,(1 - e^{-2\gamma t})]}\right\}, \tag{45}$$

where τ is the scaling time such that

$$\tau = \frac{\epsilon b_q}{2\gamma}\,(e^{2\gamma t} - 1). \tag{46}$$

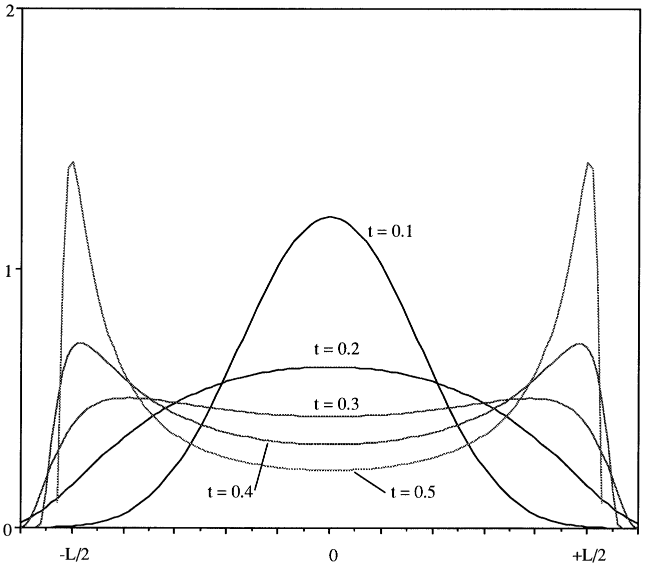


Fig. 10 Behavior of the scaling solution.

Solution (45) is called the scaling solution.

This scaling solution shows an interesting behavior. Figure 10 shows a temporal evolution of the scaling solution, where we set $L = 2$, $\gamma = g = 1$, and $\epsilon = b_{\rm q} = 1$ for simplicity. Initially, the solution has a single peak, and afterward it splits into double peaks in accordance with increasing t. These peaks signify the very long and the almost empty states of queue length. Note that we can recognize that the double-peak state corresponds to the final state of Fig. 5 for QDC. Therefore, the emerging double peak means that the deteriorating process of throughput is caused by clustering of the busy nodes and their spread.

Here, we state a supplemental comment about the above interpretation. Note that the final state in Fig. 10 does not mean that each node becomes busy or idle alternatively in the final state of QDC in Fig. 5. The final state in Fig. 10 means that state of the network becomes one of the following states:

- all nodes in the network are congested, or
- all nodes in the network are idle.

Since our network model is closed and the number of cells in the network is constant, it is impossible for the state of the network to become one of the above two states. If we consider a very large closed network and we focus on a part of the network as an open network, that partial network must become one of the above states. We should therefore interpret Fig. 10 as a clustering process of busy state nodes.

5.4 Anomalous Fluctuation and Behavior of Throughput

We are interested in the time the throughput of the network deteriorates. This is interpreted as the time when double peaks of the solution emerge. Let the variance of the solution be, initially, $\mathcal{O}(\epsilon)$. However, the variance becomes anomalously large and reaches $\mathcal{O}(1)$ as t increases. This is called the Anomalous Fluctuation or Anomalous Enhancement of Fluctuation, and often appears in state transition from unstable equilibrium to stable equilibrium. Temporal evolution (19) implies, as shown in [8], [9],

$$v(t) \propto \frac{\dot{y}(t)}{\dot{y}(0)}, \tag{47}$$

where $\dot{y}(t)$ denotes the time derivative of $y(t)$. Note that $y(t)$ and $v(t)$ describe not the throughput but the queue length in this QDC case. However, since the total number of cells in the network is constant, the variance $v(t)$ is the same as that of the throughput, and $\dot{y}(t)$ is the negative value. Therefore, we can also say that (47) is reasonable for describing the throughput.

Actually, the QDC case shown in Fig. 4 shows the time when the maximum variance corresponds to the steepest change in the average throughput.

The time the throughput of the network deteriorates corresponds to the time the variance becomes large and reaches $\mathcal{O}(1)$. The physical meaning of this is that a small fluctuation grows and causes the average value to change. The time for it, t_0, is called on-set time [8], and is obtained as

$$t_0 \simeq \frac{1}{2\gamma} \log \frac{2\gamma^2}{\epsilon b_q g}. \tag{48}$$

6. Conclusion

We have described two flow control schemes, RDC and QDC, in high-speed and large-scale networks as autonomous distributed systems. In both schemes, each node in the network handles its local traffic flow only on the basis of the information it knows. It is preferable, however, that the decision making of each node leads to high performance of the whole network. To this end, the relationship between local decision making and global performance of flow control is the essential object.

We are especially interested in the behavior of the local congestion, whether it grows and causes deterioration of the whole network performance through interaction among nodes, or it remains a local phenomenon and diminishes with time. Experimental results show interesting behaviors of throughput for both schemes. Performance of RDC is stable, but that of QDC is unstable. For QDC, the congested nodes are clustered.

To describe the behaviors, we proposed phenomenological models based on the Ω-expansion technique and gave physical interpretations. Behavior of throughput for RDC is described by a linear relaxation model of the Fokker-Planck equation. The solution of the equation is a stable Gaussian distribution. For QDC, we have assumed that the behavior of queue length can be described by a nonlinear Fokker-Planck equation, and have applied the simplest nonlinear relaxation model. The solution of the equation describes the clustering process of the congested nodes.

Main residual issues are listed as follows:

- To verify the minimal non-linear model Eq. (42). (Is Eq. (42) valid for a large χ?)
- To obtain the relationship between the round-trip time, D, and coefficients in Eq. (44).
- To obtain necessary and/or sufficient conditions which enable us to design a stable flow control scheme in high-speed networks.

Acknowledgement

The authors would like to thank anonymous reviewers for their valuable comments and instructions.

References

[1] M. Aida and K. Horikawa, “A study on stability analysis for high-speed networks based on statistical physics,” Proc. Symp. on Performance Models for Information Communication Networks, pp.1–7, Jan. 1997.

[2] F. Baskett, K.M. Chandy, R.R. Muntz, and F.G. Palacios, “Open, closed, and mixed networks of queues with different classes of customers,” Jour. Ass. Computing Machinery, vol.22, pp.248–260, 1975.

[3] K. Horikawa, M. Aida, and T. Sugawara, “Traffic control scheme under the communication delay of high-speed networks,” Proc. 1996 Int. Conf. on Multi-Agent Systems (ICMAS’96), pp.111–117, 1996.

[4] N.G. van Kampen, “A power series expansion of the master equation,” Can. J. Phys., vol.39, pp.551–567, 1961.

[5] R. Kubo, K. Matsuo, and K. Kitahara, “Fluctuation and relaxation of macrovariables,” Jour. Stat. Phys., vol.9, pp.51–96, 1973.

[6] D. Mitra, “Optimal design of windows for high speed data networks,” Proc. INFOCOM ’90, pp.1156–1163, 1990.

[7] H. Risken, Fokker-Planck Equation: Methods of Solution and Applications, Springer-Verlag, Berlin, 1989.

[8] M. Suzuki, “Passage from an initial unstable state to a final stable state,” Adv. Chem. Phys., vol.46, pp.195–278, 1981.

[9] M. Suzuki, “Anomalous fluctuation and relaxation in unstable systems,” Jour. Stat. Phys., vol.16, pp.477–503, 1976.

Masaki Aida received the B.S. and M.S. degrees in Theoretical Physics from St. Paul's University, Tokyo, Japan, in 1987 and 1989, respectively, and received the Ph.D. degree in Telecommunication Engineering from the University of Tokyo, Japan, in 1999. Since he joined NTT Laboratories in 1989, he had been mainly engaged in research on traffic issues in ATM networks and computer communication networks until March 1998. He is currently a manager at Traffic Research Center, NTT Advanced Technology Corporation (NTT-AT). His current interests include traffic issues in communication systems. He received the Young Engineer Award of IEICE in 1996. Dr. Aida is a member of the Operations Research Society of Japan.

Kenji Horikawa received the B.S. and M.S. degrees from Tokyo Institute of Technology in 1992 and 1994, respectively. In 1994 he joined NTT (Nippon Telegraph and Telephone Corporation). He has been engaged in research and development of computer networks.