PAPER

Autonomous Decentralized Flow Control in High-Speed Networks with Inhomogeneous Configurations

Chisa TAKANO[†a)], **Masaki AIDA**[††b)], *and* **Shin-ichi KURIBAYASHI**[†††c)], *Members*

SUMMARY Recent growth in computer communications has led to an increased requirement for high-speed backbone networks. In such high-speed networks, the principle adopted for a time-sensitive flow control mechanism should be that of autonomous decentralized control. In this mechanism, each node in a network manages its local traffic flow only on the basis of the local information directly available to it, although it is desirable that the individual decisions made at each node lead to high performance of the network as a whole. In our previous studies, we have investigated the behavior of local packet flows and the global performance achieved when a node is congested, and proposed the diffusion-type flow control model. However, since we used a simple and homogeneous network model in the evaluation, the results cannot be generalized. In this paper, we propose an extension of the diffusion-type flow control model in order to apply it to networks with inhomogeneous configurations. We show simulation results for two cases: different propagation delays and multiple bottlenecks. Both results show that the proposed diffusion-type flow control achieves high and stable performance even if the network is congested.
key words: *autonomous decentralized system, flow control, diffusion, feedback*

1. Introduction

In recent years, the increase in traffic transferred through the network has been further augmented by the spread of broadband access and broadband applications. To convey such a large amount of traffic, a very high-speed backbone network is required. In high-speed networks, such as gigabit or terabit/second networks, propagation delay becomes the dominant factor in the transmission delay because the speed of light provides an absolute constraint. Therefore, at any given time, a large amount of data is in the state of being propagated on links in the network (Fig. 1). The amount of such data is characterized by the bandwidth-delay product, i.e., the propagation distance multiplied by the transmission rate. Therefore, in high-speed and/or long-distance transmission, there is more data in transit on the links than there is waiting in the nodes.

Since the performance of modes is strongly dependent on their throughput capacity, but the propagation delay is constant, it is impossible to impose time-sensitive control based on collecting global information about the network. If we allow sufficient time to collect global information, the data so gathered is too old to apply to time-sensitive control. So, in a high-speed network, the principles adopted for time-sensitive control are inevitably those of autonomous decentralized systems [1]–[5].

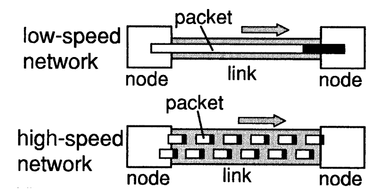

Fig. 1 Effect of large bandwidth-delay product.

This paper focuses on flow control implemented as an autonomous decentralized system. In our model, each individual network node manages the handling of its local traffic flow itself, based only on the information directly available to it. Since time-sensitive control in high-speed networks cannot collect global information about the network, nodes can use only restricted local information. We assume that it is only possible for each node to be aware of the following information: the distance between the node and adjacent nodes, the number of packets stored in the node at the present moment, and the feedback information that is received from the adjacent nodes. It is, of course, desirable that the decisions made at each node should lead to high performance of the whole network. In flow control, we use the total throughput of a network as a global performance measure.

In our previous studies, we investigated the behavior of local packet flows and the global performance measure when a node is congested, and demonstrated an appropriate flow control model through simulation results [3], [4]. In addition, we investigated the stability and adaptability of the network performance when the capacity of a link is altered [4], [5].

However, since we used a simple and homogeneous network model with uniform link delays and a single bottleneck in the evaluation, the results cannot be generalized. In this paper, we propose an extension of the diffusion-type flow control model in order to apply it to networks with inhomogeneous configurations and to demonstrate the performance of the proposed control method by using more realistic network models in the evaluation.

Manuscript received March 24, 2003.
Manuscript revised December 5, 2003.
[†]The author is with Traffic Engineering Division, NTT Advanced Technology Corporation (NTT-AT), Musashino-shi, 180-0006 Japan.
[††]The author is with NTT Information Sharing Platform Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
[†††]The author is with Seikei University, Musashino-shi, 180-8633 Japan.
a) E-mail: chisa@mitaka.ntt-at.co.jp
b) E-mail: aida.masaki@lab.ntt.co.jp
c) E-mail: kuribayashi@apm.seikei.ac.jp

The remainder of this paper is organized as follows. In Sect. 2, we discuss related studies, comparing them with our work by categorizing the flow control mechanisms. In Sect. 3, we describe the framework of our flow control method, including the measure of performance, the node model and the packet flow, before explaining our flow control mechanism. In Sect. 4, we propose an extension of the diffusion-type flow control mechanism, to make it applicable to links with different delays. This mechanism is applied to the framework outlined in Sect. 3. In Sect. 5, we describe the performance of the proposed flow control mechanism for network models with different link delays and a single bottleneck. Then, in Sect. 6, we describe the performance of the proposed flow control mechanism for network models with multiple bottlenecks in a situation where the network is shared by multiple flows. Finally, Sect. 7 provides a conclusion to this paper.

Table 1 Classification of flow control mechanisms with respect to collecting information.

	low-speed NW	high-speed NW
centralized control	1-A	1-B
decentralized control	1-C	1-D

Table 2 Classification of decentralized flow control mechanisms with respect to control delay requirement.

	decision-making	
	long time-scale	short time-scale
controlled by end hosts (end-to-end, end-to-node)	2-A	2-B
controlled by nodes (node-by-node)	2-C	2-D

2. Related Studies

In general, the technique used for flow control for high-speed networks should satisfy the following requirements:

1. It must be possible to collect the information required for used in the control mechanism.
2. The required control must be applied with minimum delay, so that it is virtually instantaneous.

As explained earlier, in high-speed networks, we cannot collect global information about the network. So, we may classify the flow control mechanisms with respect to the collecting of information as shown in Table 1.

Centralized control requires the collection of global information about the network, but this is impossible in high-speed networks. Therefore, class 1-B control mechanism cannot be realized. In low speed networks, both classes 1-A and 1-C are possible, and there are many papers which consider these classes. They mainly relate to optimization of flow control problems in the framework of solving linear programs. Techniques for addressing rate control, bandwidth assignment, deadlock resolution, resource allocation or flow fairness problems for each source by optimizing some end-to-end utility function have been reported in [6]–[10]. Reference [11] considers the issue of reducing the convergence time in solving optimization problems.

Our target is flow control in a high-speed network, for which the framework is inevitably autonomous decentralized control, so we are focusing on class 1-D.

We can also classify the decentralized flow control mechanisms from the point-of-view of the control delay requirement, as shown in Table 2.

Flow control by end hosts including TCP is widely used in current networks, and there is much research in this area.

Based on the optimization problem, [12] introduces a decentralized marking mechanism for early notification of congestion. For window-based flow control, the optimization problems of some aggregated utility functions have been explored in [13]. Solving these optimization problems, however, requires enough time to be available for calculation, and so it is difficult to apply them to decision-making in a very short time-scale.

Johari and Tan [14] studied stable end-to-end congestion control when the propagation delay is large relative to the queueing delay. Each end system requires knowledge only of its own round-trip delay. [15] reports an investigation of rate control and window control with feedback from nodes to end hosts. Each feedback message indicates the state of the buffer at the node—whether it is above or below a threshold. Since control by end hosts cannot be applied to decision-making in a time-scale shorter than the round-trip delay, it is not satisfactory for application to decision-making in a very short time-scale. Therefore, these methods are categorized as class 2-A. In high-speed networks, due to the fact that many packets are influenced by control delay, a very short control delay is required, and so class 2-B cannot be realized. Our target is node-by-node control and is categorized as class 2-D.

Bartal et al. [6] studied global optimization of flow control using local information. The motivation for their work was to enable the distributed routers in high-speed networks to make decisions on flow control as rapidly as possible, and they studied the problem in the framework of solving linear programs using distributed agents. Though this motivation is similar to that of our work, their study assumes the distributed agents can obtain detailed information about networks if we allow them to spend sufficient time. As stated above, our standpoint is based on the fact that we cannot obtain detailed, useful and up-to-date information about the whole network in a high-speed network environment, even if we do not limit the time taken to collect data.

In our previous studies, we investigated the characteristics of autonomous decentralized flow control in a high-speed network [1]–[5]. We proposed a simple and effective method of flow control in [3]–[5]. Since the proposed control method uses less information than the control methods described in [1], [2], it is relatively simple.

3. Preliminary Description of Flow Control

We assume that a target flow has a static route. In the case of Internet-based networks, to guarantee the end-to-end Quality of Service (QoS) of a flow, QoS sensitive flows use a static route (e.g. RSVP). In addition, we assume all routers in the network can employ per-flow queueing for the all target flows.

3.1 Performance Measure

Each packet in a network is either in a node or on a link. Since the packets currently stored in nodes are not being transmitted over the network, it is natural to define the total throughput of the network as a global performance measure as follows. We define the total throughput of a network at time t as the amount of data being propagated on the network [1]–[5], [16], that is the total number of packets being propagated on all links in the network at time t.

On the other hand, the only packets we can control are those stored in nodes, and not those being propagated. Thus, high performance of the whole network involves many uncontrollable packets being propagated on links. Therefore, inappropriate flow control cannot produce a condition characterized by both high performance and stability.

3.2 Node Model

Figure 2 shows the interaction of our flow control method between nodes using a network model with a 1-dimensional configuration. Although this model looks simple, it illustrates how a part of a network may be extracted from the whole network along the route of the specific flow.

All nodes have two incoming and two outgoing links, for a one-way packet stream and for feedback information, that is, node i ($i = 1, 2, \dots$) transfers packets to node $i+1$ and node $i+1$ sends feedback information (node information) to node i. For simplicity, we assume that packets have a fixed length in bits.

All nodes are capable of receiving node information from, and sending it to, adjacent downstream and upstream nodes, respectively. Each node i can receive node information sent from the downstream node $i+1$, and can send the node information about node i itself to the upstream node $i-1$.

When node i receives node information from downstream node $i+1$, it determines the transmission rate for packets to the downstream node $i+1$ using the received node information and adjusts its transmission rate towards the downstream node $i+1$ accordingly. The framework of node behavior and flow control is summarized as follows:

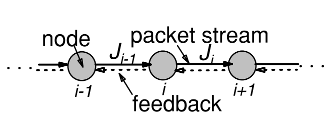

Fig. 2 Interaction between nodes.

- Each node i autonomously determines the transmission rate J_i based only on the information available to it, that is, the node information obtained from the downstream node $i+1$ and its own node information.
- The rule for determining the transmission rate is the same for all nodes.
- Each node i adjusts its transmission rate towards the downstream node $i+1$ to J_i.
  (If there are no packets in node i, the packet transmission rate is 0.)
- Each node i autonomously creates node information according to a predefined rule and sends it to the upstream node $i-1$.
- The rule for creating the node information is the same for all nodes.
- Packets and node information both experience the same propagation delay.

As mentioned above, the framework of our flow control model involves both autonomous decision-making by each node and interaction between adjacent nodes. There is no centralized control mechanism in the network, since, as described earlier, it is impossible to realize centralized control in a high-speed network environment.

3.3 Packet Flow

In this paper, we focus on the stability of flow control in the congested state, and consider packet flow in a heavy-traffic environment. The packet flow is defined as the number of packets sent per unit of time, and in a heavy-traffic environment it is the same as the transmission rate towards the downstream node. That is, if the transmission rate specified by node i is $J_i(t)$, we let the packet flow be $J_i(t)$. This is because node i has sufficient packets to transfer. Hereafter, we identify the packet flow with the transmission rate specified by the node.

The packet flow $J_i(t)$ should be controlled by the behavior of node i in the framework described in Sect. 3.2. This means that the packet flow from a node can be expressed using the node information obtained from the downstream node $i+1$ and information its about the node itself. So, we consider the packet flow determined through a certain function, F, as

$$J_i(t) = F(n_i(t), d_i, r_i(t-d_i), n_{i+1}(t-d_i)), \qquad (1)$$

where $n_i(t)$ denotes the number of packets in node i at time t, d_i denotes the propagation delay between node i and node $i+1$, and $r_i(t-d_i)$ is the target transmission rate specified by the downstream node $i+1$ as node information. In addition, $r_i(t-d_i)$ and $n_{i+1}(t-d_i)$ are notified from the downstream node $i+1$ with the propagation delay d_i. The interval between two successive notifications from the downstream node $i+1$ is constant and is denoted by τ_{i+1}.

If there is no packet loss in the network, the temporal evolution of $n_i(t)$ is expressed as a continuous equation,

$$n_i(t+\epsilon) - n_i(t) = \epsilon\,[J_{i-1}(t-d_{i-1}) - J_i(t)]\,, \tag{2}$$

where $\epsilon > 0$ is a small number.

4. Flow Control Models

In this section, we propose an extension of the diffusion-type flow control described in [3]–[5], in order to apply it to networks which include different link lengths.

4.1 Diffusion-Type Flow Control

Node i's packet transmission rate to the downstream node $i+1$ is defined as

$$J_i(\alpha, t) = \alpha\, r_i(t-d_i) - D_i\,(n_{i+1}(t-d_i) - n_i(t)). \tag{3}$$

The first term on the right hand side of Eq. (3) reflects the target rate specified by the downstream node, and the second term, which is called the diffusion term, is proportional to the rate of change of the packet density. We call α (≥ 1) and D_i (> 0) the flow intensity multiplier and the diffusion coefficient, respectively. Here, we choose $D_i = D/d_i$ where D is a constant; in the previous studies, [3]–[5], D_i was a constant and independent of i.

In addition, the node information of node i sent to the upstream node $i-1$ is determined every fixed period τ_i as

$$r_{i-1}(t) = J_i(1, t). \tag{4}$$

To implement this flow control, we use the following flow control model. Since the packet flow is restricted by the link capacity L_i, the flow control is expressed as follows:

$$J_i(\alpha, t) = \min(\max(\tilde{J}_i(\alpha, t), 0), L_i), \tag{5}$$
$$r_{i-1}(t) = J_i(1, t), \tag{6}$$

where

$$\tilde{J}_i(\alpha, t) = \alpha\, r_i(t-d_i) - D_i\,(n_{i+1}(t-d_i) - n_i(t)). \tag{7}$$

In the framework of Eqs. (5)–(7), the feedback information of i passed to the upstream node $i-1$ is a pair of values ($r_{i-1}(t)$, $n_i(t)$). Feedback information is created periodically, and τ_i denotes the time interval between successive notifications of feedback information from node i to the upstream node $i-1$. We choose $\tau_i \propto d_{i-1}$; in the previous studies [3]–[5], τ_i was constant and independent of i. In the rest of this paper, we choose $\tau_i = d_{i-1}$, that is one set of feedback data information is always being transmitting on the link. Moreover, the packet flow $J_i(t)$ in node i is revised whenever feedback information arrives from the downstream node $i+1$ (with a period of $\tau_{i+1} = d_i$).

The propagation delay d_{i-1} may vary on account of a change in network topology. In this case, each node needs to be aware of the new propagation delay for the adjacent link and to update the time interval τ_i. However, since the time interval between changes in network topology is very long in comparison with the time scale of the behavior of this flow control, we consider that the propagation delay is generally constant in this evaluation. The following is a description of the flow control method, which determines the transmission rate such that the difference between the number of packets in one node and that in the downstream node becomes smaller.

To provide a rough estimate the temporal evolution, we replace i with x and apply continuous approximation. Then the propagation delay becomes $d_i \to 0$ for all i and in the limit, the packet flow (3) is expressed as

$$J(x, t) = \alpha\, r(x, t) - D\,\frac{\partial n(x, t)}{\partial x}. \tag{8}$$

The temporal evolution of the number of packets at x may be expressed as a diffusion type equation,

$$\frac{\partial n(x, t)}{\partial t} = -\alpha\,\frac{\partial r(x, t)}{\partial x} + D\,\frac{\partial^2 n(x, t)}{\partial x^2}, \tag{9}$$

by using the continuous equation,

$$\frac{\partial n(x, t)}{\partial t} + \frac{\partial J(x, t)}{\partial x} = 0. \tag{10}$$

Equation (9) is a sort of diffusion equation. That is, our method aims to perform flow control using the analogy of a diffusion phenomenon. We can expect packets in a congested node to be redistributed over the whole network and normal network conditions to be restored after some time.

Note that, to obtain the limit (8) from Eq. (3), it is necessary to choose

$$D_i \propto \frac{1}{d_i}, \tag{11}$$

and the flow intensity multiplier α is independent of d_i.

4.2 Drift-Type Flow Control

If we cannot use the information of $n_{i+1}(t-d_i)$ specified by the downstream node, diffusion-type flow control cannot determine the packet flow. So, we consider a flow control model in which $n_{i+1}(t-d_i)$ in Eq. (1) is replaced with the constant threshold n_s (≥ 0). Since the packet flow is restricted by the link capacity L_i, this flow control is expressed as follows:

$$\tilde{J}_i(\alpha, t) = \alpha\, r_i(t-d_i) - D_i\,(n_s - n_i(t)), \tag{12}$$
$$n_s = \text{constant}\ \ (\geq 0). \tag{13}$$

5. Evaluation for Network with a Single Bottleneck

In this section, we use a simulation to show the importance of feedback information and the importance of selecting an appropriate diffusion coefficient. In our simulation study, we consider the case where the capacity of a link in the network is suddenly reduced to a narrow bandwidth. This situation

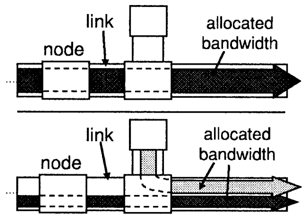

Fig. 3 Example of a bottleneck link.

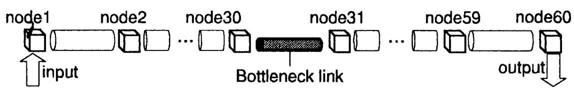

Fig. 4 Network model with a bottleneck link.

occurs when, for example, there is fault on a link, or background traffic occupies bandwidth. For example, Fig. 3 illustrates a change in available bandwidth influenced by traffic in a different flow. In high-speed networks, no node is directly aware of the change in the state of the link resulting in a new capacity.

First, we compare network performance for the two different flow control mechanisms described in Sect. 4 by using a simple and homogeneous network model having a single bottleneck and equal propagation delays between adjacent nodes. We show that the diffusion-type flow control mechanism provides high network performance irrespective of the capacity of the bottleneck link.

Next, we consider network models with different propagation delays between the adjacent nodes and investigate the performance of the diffusion-type flow control using these models. We show that the diffusion-type flow control absorbs the complexity of the network model, and achieves high performance and stability even if the link delays in the network are different.

5.1 Network Model and Simulation Conditions

Figure 4 shows our network model, which is an open network with a 1-dimensional configuration. The network has a bottleneck link and a corresponding congested node. All the other nodes and links are in the same condition as each other. This model simulates the situation when congestion occurs at a certain node. We are interested in the behavior of the local congestion, that is, whether:

- it causes deterioration in the total network performance through interaction among nodes, or
- it diminishes with time.

Detailed conditions of our network model are listed below.

- Number of nodes: $m = 60$. Each node specified by i ($i = 1, 2, \ldots, 60$).
- Index of the congested node: $i = 30$.
- Mean number of total packets in the network: $N = 6000$ (the packet input process is the Bernoulli process and is independent of the output process, but the input packet rate is equivalent to the output one so that the mean number of packets is kept constant).
- Bandwidth of each link except the bottleneck link: $L_i = 100$ [packets/unit time] ($i \neq 30$).
- Bandwidth of the bottleneck link (between nodes $i = 30$ and 31): $L_{30} = 25$, 50, or 75 [packets/unit time] (that is, 1/4, 1/2, or 3/4 of the bandwidth of other links).

To investigate the stability under congestion, in addition to the above conditions, we set the initial condition for congested node $i = 30$ as follows.

- Number of packets in node $i = 30$ at time $t = 0$: 400.
- The other 5600 packets are randomly configured in other nodes and on other links.
- Propagation delays between adjacent nodes (three models):

model 1: All distances between adjacent nodes are the same: 1 (unit time).

model 2: The distance between any node and the adjacent node can be one of two values, short or long, (that is, it has a small or large propagation delay), where the ratio of the length of the short links to that of the long links is 1:50. Each link value appears with the probability 1/2. So, we choose about 0.04 or 1.96 for the propagation delays of the 59 links so that the mean propagation delay is 1.0 and the variance is 0.94.

model 3: The length of each link is determined according to a log-normal distribution, where the mean delay in the network model is 1.0 and the variance is 5.2.

The values of the parameters used in our flow control models for the evaluation are as follows:

$$(D_i, \tau_i) = \begin{cases} (0.1 \times \frac{1}{d_i}, d_{i-1}), \\ (0.1, 1), \\ (0.1 \times d_i, \frac{1}{d_{i-1}}), \end{cases} \tag{14}$$

$$\alpha_i = 1.01, \tag{15}$$

$$n_s = 0, \text{ or } 60 \text{ (for drift-type control)}. \tag{16}$$

The first combination of values for (D_i, τ_i) in Eq. (14) corresponds to our flow control as described in Sect. 4. The second one corresponds to the flow control shown in our previous studies [3]–[5]. In addition, we consider the third combination in order to compare it with the other two. Note that these three pairs of parameters provide the same control mechanism in a network model having the equal link delays.

5.2 Simulation Results for Model 1: Total Throughput and Stability

We compare the total throughput of the network for the diffusion and drift type flow control models described in the previous section by using *model 1*, and discuss the stability of the flow control models in a high-speed network environment. Note that the three variations of Eq. (14) for the

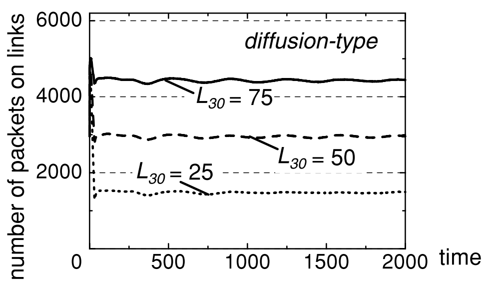

Fig. 5 Total throughput of the network in *model 1* (diffusion-type).

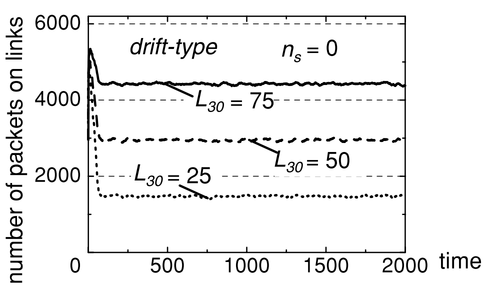

Fig. 6 Total throughput of the network in *model 1* (drift-type $n_s = 0$).

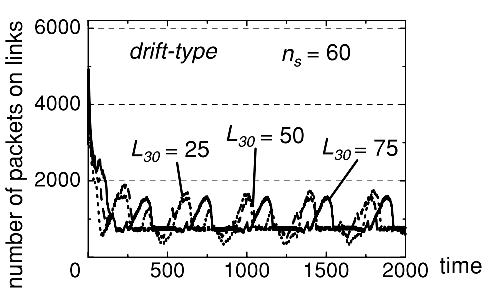

Fig. 7 Total throughput of the network in *model 1* (drift-type $n_s = 60$).

diffusion-type flow control are the same in this evaluation ($d_i = 1$ (constant)).

Figures 5, 6 and 7 show the results of the total throughput for the diffusion-type and two variations of the drift-type ($n_s = 0$ and 60) flow control models, respectively. The horizontal axis denotes the simulation time and the vertical axis the total throughput (i.e., the total number of packet being propagated on links). The three lines in these figures shows the results for the cases where the capacity of the bottleneck link $L_{30} = 25$, 50, and 75.

We discuss the results from a quantitative point of view. Figure 5 shows that the total throughput decreases with time but becomes stable at around 1500, 3000, or 4500 corresponding to a capacity of the bottleneck link $L_{30} = 25$, 50, or 75, respectively. For a given capacity of the bottleneck link L_{30}, the maximum value of the sustainable total throughput (the maximum number of packets being propagated stably on links) should be L_{30} packets/link × 59 links. Thus, the diffusion-type flow control achieves about 100% of the maximum value of the total throughput (in other words, transfer efficiency is about 100%) irrespective of the value of L_{30}. In the diffusion-type control model, although no node is aware of the bandwidth of the bottleneck link, stable and high performance is achieved.

Figure 6 shows that for drift-type flow control with the value of $n_s = 0$, the total throughput is also high and stable irrespective of the value of L_{30}. In this case also, the drift-type flow control provides about 100% transfer efficiency. In Fig. 7, on the other hand, when n_s is large such as $n_s = 60$, the values of total throughput are in all cases small and unstable.

Next, we investigate the distribution of packets in the individual nodes for these flow control models. Figures 8–10 show the simulation result for flow control models when the capacity of the bottleneck link L_{30} is 50. The horizontal axis of each graph denotes node ID and the vertical axis the number of packets stored in the node, i.e., the queue length in the node. In addition, t denotes the simulation time and initially $t = 0$.

As mentioned above, both the diffusion-type and the drift-type control with the value of $n_s = 0$ provide stable and high total throughput. However, Fig. 9 shows that for the drift-type flow control with the value of $n_s = 0$, there is a drastic increase in the queue length of the congested node $i = 30$, while for the diffusion-type flow control, the distribution of the number of packets stored in nodes is distributed smoothly over the network with time. Since a large queue length causes packet loss, the drift-type flow control with the value of $n_s = 0$ is not appropriate. For drift-type control with the value of $n_s = 60$, the number of packets in the congested node $i = 30$ decreases with time, but the distribution of the number of packets is uneven and the maximum number of packets in a node is very large.

If we can choose an appropriate value for the threshold n_s for the capacity of the bottleneck link L_{30}, the total throughput may be stable and adaptive for drift-type flow control. However, drift-type control cannot achieve high performance, since nodes cannot be aware of information about the bandwidth of the bottleneck link in a high-speed network environment.

We can recognize from the above results that $n_{i+1}(t-d_i)$ used in the diffusion-type flow control is essential for providing the superior performance from the point of view of transfer efficiency, stability, and the avoidance of congestion.

Hereafter, focusing on the diffusion-type flow control, we will evaluate the network performance when the network model is more complicated.

5.3 Simulation Results for Models 2 and 3: Total Throughput and Stability

We discuss the performance and stability of the diffusion-type flow control model for *model 2* and *model 3*, through evaluation of the total throughput.

In this discussion, the value of the diffusion coefficient D_i shown in Sect. 4.1 is justified.

Figure 11 shows the total throughput for *model 2* and

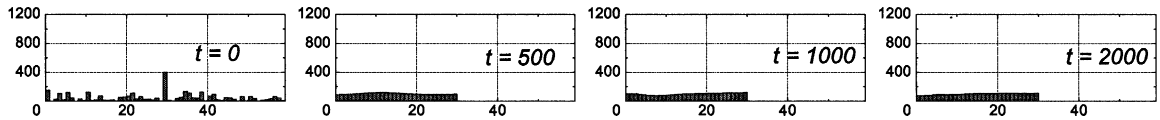

Fig. 8 Temporal evolution of the number of packets in each node for diffusion-type flow control in *model 1*.

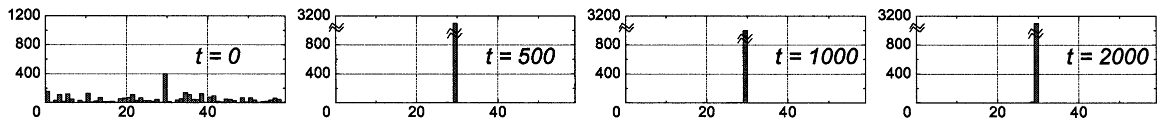

Fig. 9 Temporal evolution of the number of packets in each node for drift-type flow control (n_s = 0) in *model 1*.

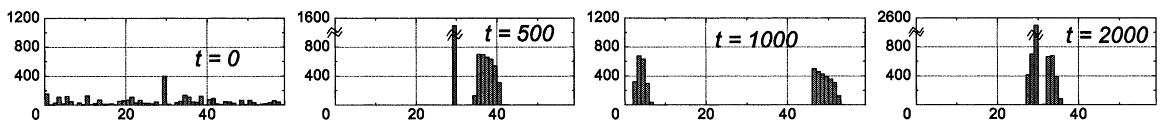

Fig. 10 Temporal evolution of the number of packets in each node for drift-type flow control (n_s = 60) in *model 1*.

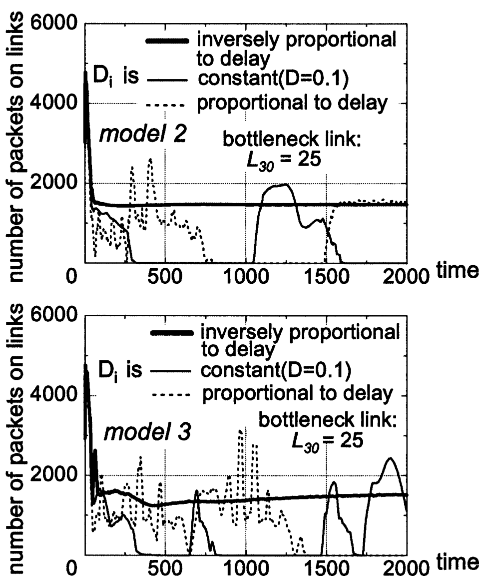

Fig. 11 Total throughput of the network (*model 2* and *model 3*) (in the case where the delays between the adjacent nodes are different).

model 3, where the horizontal axis denotes the simulation time and the vertical axis denotes the total throughput. The capacity of the bottleneck link $i = 30$ is $L_{30} = 25$. The three lines in this figure show the results from the diffusion-type flow control model with $D_i = 0.1/d_i$, 0.1, and $0.1 \times d_i$. We can see in Fig. 11 that the total throughput becomes stable at a higher level of performance in the cases where D_i is inversely proportional to the propagation delay, for both *model 2* and *model 3*. These results imply that Eq. (11) is appropriate for providing high performance and stability in networks with inhomogeneous configurations.

Although other combinations of (D_i, τ_i) are possible in principle, they have no physical meaning and give lower performance than the case for $D_i = 0.1/d_i$.

The values of total throughput achieved by the model with $D_i \propto 1/d_i$ shown in Fig. 11 are almost the same as those for the cases where the capacity of the bottleneck link is 25, as shown in Fig. 5. This means the appropriate setting of the diffusion coefficient Eq. (11) allows the complexity of the network model to be absorbed, and provides high performance and stability even if the configuration of the network becomes complex.

Next, we consider the temporal evolution of the distribution of packets stored in nodes, the results being shown in Figs. 12–17. The first three figures show the results for *model 2* and the last three show those for *model 3*. For each model the figures show the cases for $D_i = 0.1/d_i$, 0.1, and $0.1 \times d_i$, respectively.

In both models, the case of $D_i = 0.1/d_i$ exhibits a smooth diffusion of the packet distribution over the network.

6. Evaluation for Network with Multiple Bottlenecks

In this section, we consider the performance of the diffusion-type flow control in a situation where there are many flows sharing links on the route of the target flow.

We consider a one-dimensional network model with 60 nodes (Fig. 18). As in Fig. 4, this model shows a part of a network extracted from the whole network along the route of a specific flow. The bandwidth of all links is 100 packets/unit time, and link delays are determined by a log-normal distribution, as in Sect. 5.3. Background flows, which share the links on the route of the target flow, are generated according to a Poisson process. These generated flows select two different nodes in the network model at random and share the links between the two selected nodes. The lifetime of the background flow follows an exponential distribution. The packets generated by each background flow are

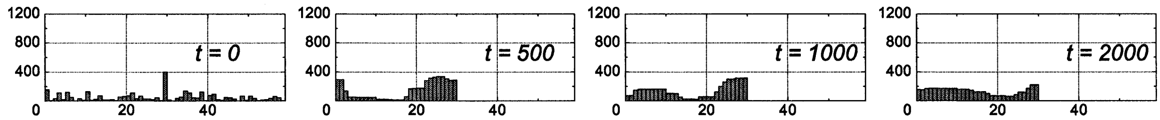

Fig. 12 Temporal evolution of the number of packets in each node for diffusion-type flow control (with $D_i = 0.1/d_i$) in *model 2*.

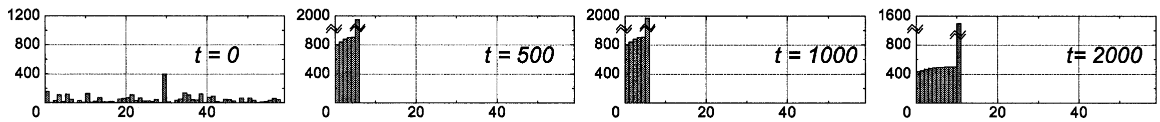

Fig. 13 Temporal evolution of the number of packets in each node for diffusion-type flow control (with $D_i = 0.1$) in *model 2*.

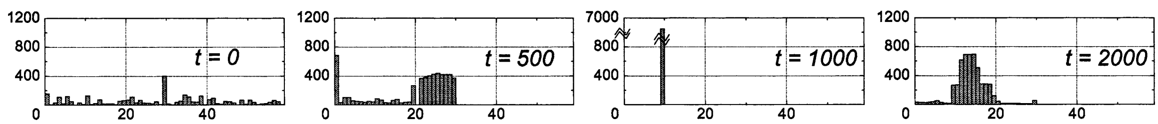

Fig. 14 Temporal evolution of the number of packets in each node for diffusion-type flow control (with $D_i = 0.1 \times d_i$) in *model 2*.

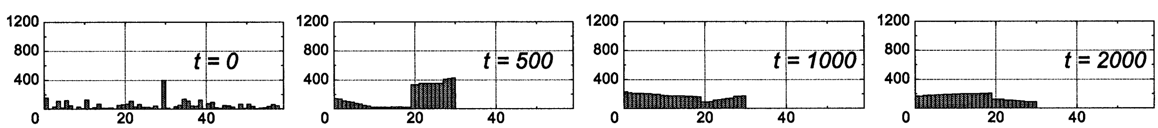

Fig. 15 Temporal evolution of the number of packets in each node for diffusion-type flow control (with $D_i = 0.1/d_i$) in *model 3*.

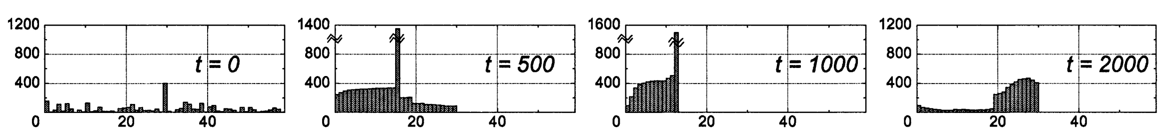

Fig. 16 Temporal evolution of the number of packets in each node for diffusion-type flow control (with $D_i = 0.1$) in *model 3*.

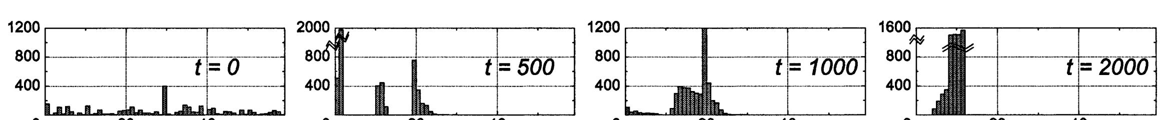

Fig. 17 Temporal evolution of the number of packets in each node for diffusion-type flow control (with $D_i = 0.1 \times d_i$) in *model 3*.

also regulated by the diffusion-type flow control. Nodes in the network employ per-flow queueing and round robin fair scheduling.

If at most N flows share a link, the maximum sustainable total throughput for the target flow is $100/N \times 59$. Figures 19 and 20 show the maximum sustainable total throughput and the total throughput obtained from simulation. The horizontal axis denotes the simulation time and the vertical axis denotes the total throughput. Figure 19 shows the case where the mean interval between the occurrence of background flows is 1000 units of time, and the mean lifetime of a flow is 2000 units of time, while Fig. 20 shows similar results with flow arrival and lifetime values of 3000 and 3000. Here, we choose $D_i = 0.1/d_i$ and $\alpha = 1.01$.

From both figures, the diffusion-type flow control can be seen to achieve high performance, but adapting to the network conditions.

7. Conclusions

This paper has presented the performance and stability of the diffusion-type flow control mechanism in the framework of an autonomous decentralized system capable of supporting high-speed networks. In this framework, nodes handle

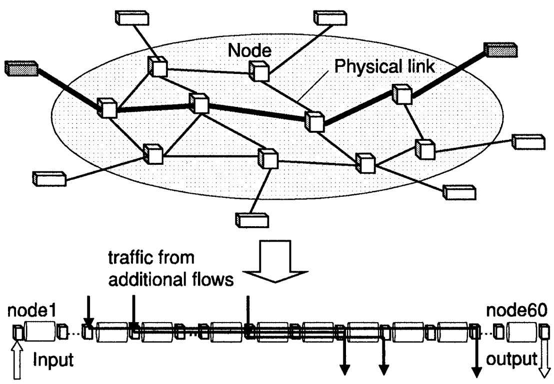

Fig. 18 Network model with multiple bottlenecks.

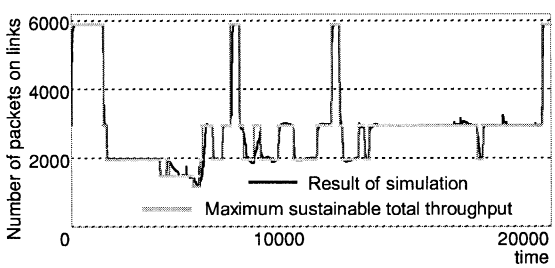

Fig. 19 Total throughput of network (mean interval of flow arrival = 1000 units of time; mean lifetime of flow = 2000 units of time).

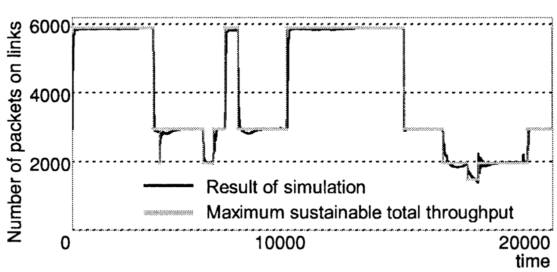

Fig. 20 Total throughput of network (mean interval of flow arrival = 3000 units of time; mean lifetime of flow = 3000 units of time).

their local traffic flows themselves based only on the information immediately available to them.

We have proposed an extension of the diffusion-type flow control model in order to apply it to networks with inhomogeneous link delays. The extension is based on an investigation of the appropriate value of the diffusion coefficient, D_i, in our flow control model, and we found the condition $D_i \propto 1/d_i$ by comparison with the diffusion equation.

We have shown simulation results for two cases: different propagation delays and multiple bottlenecks. Both results have shown that the proposed diffusion-type flow control achieves high and stable performance even when the network is congested.

In particular, if we choose the diffusion coefficient as $D_i \propto 1/d_i$, the diffusion-type flow control copes with increased complexity in the network model, and achieves high performance and stability even if the link delays in the network are different.

Finally, we briefly discuss the qualitative characteristics of the values of α and D. The previous studies [4], [5] showed that the diffusion-type flow control with $\alpha > 1$ achieves rapid recovery of the total throughput when the bottleneck link is restored. However, a larger value of α may cause instability in the total throughput. If we can choose a large value of D, the instability caused by α may be limited.

If we determine the transmission rate using Eq. (3), a larger value of D is appropriate because the corresponding diffusion equation exhibits rapid diffusion and so the packet density is reduced quickly. However, since there is the constraint on the bandwidth L_i, we determine the transmission rate by Eq. (5) instead of Eq. (3). A large value of D causes a large difference between Eq. (3) and Eq. (5). Since the transmission rate determined by Eq. (3) is essential for the diffusion effect, the difference prevents the smooth diffusion of the packet distribution.

The issue of the optimal values of α and D is for further study.

References

[1] M. Aida and K. Horikawa, "Stability analysis for global performance of flow control in high-speed networks based on statistical physics," IEICE Trans. Commun., vol.E82-B, no.12, pp.2095–2106, Dec. 1999.
[2] K. Horikawa, M. Aida, and T. Sugawara, "Traffic control scheme under the communication delay of high-speed networks," Proc. Int. Conf. on Multi-Agent Systems (ICMAS'96), pp.111–117, Dec. 1996.
[3] M. Aida and C. Takano, "Stability of autonomous decentralized flow control schemes in high-speed networks," Proc. IEEE ICDCS 2002 Workshop (ADSN 2002), pp.63–68, July 2002.
[4] C. Takano and M. Aida, "Stability and adaptability of autonomous decentralized flow control in high-speed networks," IEICE Trans. Commun., vol.E86-B, no.10, pp.2882–2890, Oct. 2003.
[5] C. Takano and M. Aida, "Stability and adaptability of autonomous decentralized flow control," Proc. 6th International Symposium on Autonomous Decentralized Systems (ISADS 2003), pp.311–318, Pisa, Italy, April 2003.
[6] Y. Bartal, J. Byers, and D. Raz. "Global optimization using local information with applications to flow control," Proc. 38th Ann. IEEE Symp. on Foundations of Computer Science (FOCS), pp.303–312, Oct. 1997.
[7] B. Awerbuch and Y. Azar, "Local optimization of global objectives: Competitive distributed deadlock resolution and resource allocation," Proc. 35th Ann. IEEE Symp. on Foundations of Computer Science, pp.240–249, 1994.
[8] D. Lapsley and S. Low, "An IP implementation of optimization flow control," Proc. IEEE GLOBECOM'98, Nov. 1998.
[9] S.H. Low and D.E. Lapsley, "Optimization flow control-I: Basic algorithm and convergence," IEEE/ACM Trans. Netw., vol.7, no.6, pp.861–874, 1999.
[10] K. Kar, S. Sarkar, and L. Tassiulas, "A simple rate control algorithm for maximizing total user utility," Proc. IEEE INFOCOM 2001, pp.133–141, 2001.
[11] B. Awerbuch and Y. Shavitt, "Converging to approximated max-min flow fairness in logarithmic time," Proc. IEEE INFOCOM'98, pp.1350–1377, March 1990.
[12] S. Kunniyur and R. Srikant, "A decentralized adaptive ECN marking algorithm," Proc. IEEE GLOBECOM'00, pp.1719–1723, 2000.
[13] J. Mo and J. Walrand, "Fair end-to-end window based congestion

control," IEEE/ACM Trans. Netw., vol.8, no.5, pp.556–567, Oct. 1999.
[14] R. Johari and D. Tan, "End-to-end congestion control for the Internet: Delays and stability," IEEE/ACM Trans. Netw., vol.9, no.6, pp.818–832, Dec. 2001.
[15] F. Bonomi, D. Mitra, and J.B. Seery, "Adaptive algorithms for feedback-based flow control in high-speed, wide-area ATM networks," IEEE J. Sel. Areas Commun., vol.13, no.7, pp.1267–1283, Sept. 1995.
[16] D. Mitra, "Optimal design of windows for high speed data networks," Proc. INFOCOM'90, pp.1156–1163, 1990.

Chisa Takano received the B.E. in Telecommunication Engineering from Osaka University, Japan, in 2000. In 2000 she joined the Traffic Research Center, NTT Advanced Technology Corporation (NTT-AT). She has been engaged in research and development of computer networks. She received IEICE's Young Investigators Award in 2003 and Information Network Research Award in 2004.

Masaki Aida received his B.Sc. and M.Sc. in Theoretical Physics from St. Paul's University, Tokyo, Japan, in 1987 and 1989, and received the Ph.D. in Telecommunication Engineering from the University of Tokyo, Japan, in 1999. Since joining NTT Laboratories in 1989, he has been mainly engaged in research on traffic issues in ATM networks and computer communication networks. From March 1998 to March 2001, he was a manager at the Traffic Research Center, NTT Advanced Technology Corporation (NTT-AT). He is currently a Senior Research Engineer at NTT Information Sharing Platform Laboratories. His current interests include traffic issues in communication systems. He received IEICE's Young Investigators Award in 1996, and Information Network Research Awards of IEICE in 2001, 2002, 2003, and 2004. Dr. Aida is a member of the IEEE and the Operations Research Society of Japan.

Shin-ichi Kuribayashi received his B.E., M.E., and Ph.D. in engineering from Tohoku University, Sendai, in 1978, 1980, and 1988 respectively. He joined the NTT Electrical Communications Labs in 1980. He has been engaged in the design and development of DDX and ISDN packet switching, ATM, PHS, and the IMT2000 and IP-VPN systems. He researched distributed communication systems at Stanford University from December 1988 through December 1989. He participated in international standardization on the ATM signaling and IMT2000 signaling protocols at the ITU-T SG11 from 1990 through 2000. He received Information Network Research Award in 2004. Since April 2004, he has been a Professor in Department of Applied Physics, Faculty of Engineering, Seikei University.