Worst-Case Delay Control in Multigroup Overlay Networks

Wanqing Tu, *Member*, *IEEE*, Cormac J. Sreenan, *Member*, *IEEE*, and Weijia Jia, *Member*, *IEEE*

Abstract—This paper proposes a novel and simple *adaptive control algorithm* for the effective delay control and resource utilization of end host multicast (EMcast) when the traffic load becomes heavy in a multigroup network with real-time flows constrained by (σ, ρ) regulators. The control algorithm is implemented at the overlay networks and provides more regulations through a novel (σ, ρ, λ) regulator at each group end host who suffers from heavy input traffic. To our knowledge, it is the first work to incorporate traffic regulators into the end host multicast to control heavy traffic output. Our further contributions include a theoretical analysis and a set of results. We prove the existence and calculate the value of the rate threshold ρ^* such that for a given set of K groups, when the average rate of traffic entering the group end hosts $\bar{\rho} > \rho^*$, the ratio of the worst-case multicast delay bound of the proposed (σ, ρ, λ) regulator over the traditional (σ, ρ) regulator is $O(\frac{1}{K^n})$ for any integer n. We also prove the efficiency of the novel algorithm and regulator in decreasing worst-case delays by conducting computer simulations.

Index Terms—Worst-case delay control, overlay multicast, multiple groups, traffic control.

1 Introduction

End host multicast (EMcast) has emerged as an alternative to interdomain Internet Protocol (IP) multicast. A large number of EMcast protocols [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] have been proposed since Narada [1] demonstrated the feasibility of EMcast. Few of these protocols were designed for multigroup networks. In a multigroup network, end hosts may join in several multicast groups. When one end host belongs to more than one group, the end host has to process multiple simultaneously entering flows. As such, and because the group flows are usually high-rate real-time flows, the end hosts that join in multiple groups are prone to become bottlenecks, incurring unacceptable multicast delays and compromised scalability performance.

A popular way to free bottlenecks is to design capacity-aware EMcast protocols [5], [12], [13] that assign the direct child members for each end host based on the end host output capacity. Thus, the end host has enough capacity to output the received packets to all its direct child members and will not become a communication bottleneck. However, such bottleneck avoidance is achieved at the cost of increasing the lengths of the multicast paths from the source to the group receivers. As illustrated in Fig. 1, suppose each flow in the multicast network has the uniform rate ρ and each end host has the same output capacity $C = 5\rho$. Fig. 1a gives the capacity-aware tree when all the end hosts join in exactly one group, in which only one transmission flow exists. In this case, each end host may have at most $\left\lfloor \frac{5\rho}{\rho} \right\rfloor = 5$ direct child members. Therefore, end host 0 (where the flow enters) has the capacity to output packets to all other end hosts 1, 2, 3, and 4 simultaneously. When the end hosts subscribe to two single-source groups, however, they may only connect to at most $\lfloor \frac{5\rho}{2\rho} \rfloor = 2$ child members directly. The reconstructed multicast tree is shown in Fig. 1b. End host 0 will not forward the packets to end hosts 3 and 4, who will receive the packets from end host 1 instead. It can be seen that the height of the multicast tree increases with the number of end host groups. Therefore, longer multicast delays are created. Such longer multicast delays are not only caused by the propagation and transmission delays of the newly added underlying links, but also by the way the packets transmit in EMcast. In EMcast, the packets are forwarded by the end hosts and, therefore, experience delays when they transmit between the IP layer and the application layer (we analyzed such delays in [14]). Moreover, end hosts usually take more time to replicate and forward packets than network routers because of the end hosts' lesser capacities (for example, CPU clock speed). Under a heavy network traffic load, network transmission delays are already long. If path lengths are increased, then an unacceptable delay performance usually results. Hence, instead of the capacity-aware scheme, a multicast traffic control mechanism that does not increase the tree height is desirable.

There are two classical traffic control methods: the leaky-bucket mechanism [20], [21], [22] and the (σ, ρ) regulator. The leaky-bucket mechanism enforces a rigid output pattern at the average rate, irrespective of the burstiness

- W. Tu and C.J. Screenan are with the Department of Computer Science, University College Cork, Western Road, Cork, Ireland. E-mail: {wt1, cjs}@cs.ucc.ie.
- W. Jia is with the Department of Computer Science, City University of Hong Kong, Hong Kong, P.R. China. E-mail: ijia@cs.cityu.edu.hk.

Manuscript received 24 May 2006; revised 16 Nov. 2006; accepted 18 Jan. 2007; published online 2 Feb. 2007.
Recommended for acceptance by P. Mohapatra.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-0134-0506.
Digital Object Identifier no. 10.1109/TPDS.2007.1074.

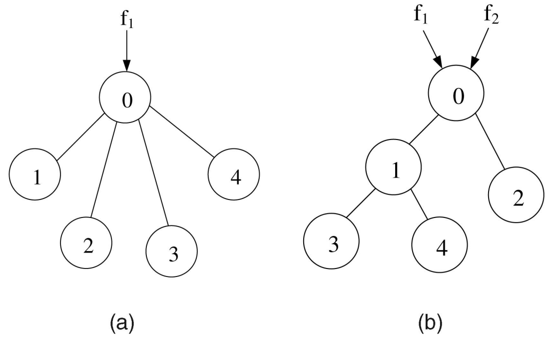


Fig. 1. An example of constructing the capacity-aware multicast tree. (a) For one single-source group and (b) for two single-source groups with the capacity-aware scheme.

of input traffic. For real-time applications, a more flexible mechanism is needed, allowing the processing of bursty flows within short delays and, preferably, with no data loss. The (σ, ρ) regulator is such a mechanism that introduces burstiness into the traffic model. The burstiness constraints that the regulator considers for a given traffic stream partially characterize the stream in the following way. Given any positive number ρ, there exists a (possibly infinite) number σ such that if the traffic is fed to a server that works at rate ρ while there is work to be done, then the size of the backlog will never be larger than σ [15], [16] (we explain the physical meaning of σ and ρ in Section 3). Our motivation in this paper is to decrease the worst-case delay bound (WDB) in multigroup EMcast networks by adopting a new algorithm to control heavy traffic. We employ the (σ, ρ) regulator as the model to analyze the WDBs of real-time flows. By the WDB, we refer to the longest packet delay at the end host who is the last one in the group to receive the packets. Like tree stability and link stress, the WDB is an important metric for EMcast. The WDB indicates whether all of the communication groups can achieve an acceptable delay performance (that is, the performance that meets the end-to-end delay bound requirements) or not. The decision to allow a new group to join the network is therefore based on the WDB. A shorter WDB improves the network's ability to host more groups.

We propose a novel and simple *adaptive control algorithm* that is implemented in the overlay network. Unlike capacity-aware EMcast protocols, our algorithm adaptively employs the novel (σ, ρ, λ) regulators to free bottlenecks without increasing the lengths of multicast paths (λ is a control parameter that will be introduced in Section 3). With the proposed regulator, when the network traffic becomes heavy, the forwarding of flows at each end host is controlled in turn based on the current network state. To our knowledge, it is the first work to incorporate traffic regulators into EMcast. Aside from the *adaptive control algorithm*, we present a theoretical analysis and a set of results on the WDB for a single regulated end host and a regulated EMcast network, respectively. Denote the WDBs of the real-time flows constrained by the (σ, ρ, λ) and the (σ, ρ) regulators as $\hat{D}$ and D, respectively, the average input rate of the real-time flows as $\bar{\rho}$, and the end host's available output capacity as C. To be specific, our contributions include the following:

- The existence of the rate threshold ρ^* is proved such that $\hat{D} \geq D$ when $\bar{\rho} \leq \rho^*$ and $\hat{D} \leq D$ when $\bar{\rho} \geq \rho^*$.
- For a single regulated end host with K input flows, $\rho^* = 0.73C$ ($\rho^* = 0.79C$) for the homogeneous (heterogeneous) flows, and the ratio of $\hat{D}$ over D when $\bar{\rho} > \rho^*$ is $O(\frac{1}{K^n})$ for both homogeneous and heterogeneous flows, where n is any positive integer.
- For a multicast group G with the size n, the height of dynamic shared cluster tree (DSCT) EMcast tree [14] is upper bounded by $\lceil \log_k^{[k+(n-j_1)(k-1)]} \rceil$, where k (set as 3 in [11]) is a random positive integer decided by the group size and the application requirements, and $j_1 \in [0, k-1]$.
- For a multigroup network with K groups that are denoted as $G^I (I \in [1, K])$, if each group G^I has n_I members that construct a DSCT tree, then we may derive $\rho^* = 0.73C$ ($\rho^* = 0.79C$) for the homogeneous (heterogeneous) flows in the multigroup network. The ratio of $\hat{D}$ over D when $\bar{\rho} > \rho^*$ is $O(\frac{1}{K^n})$ for both homogeneous and heterogeneous flows, where n is any positive integer.

The rest of the paper is organized as follows: Section 2 introduces the related work. Section 3 gives the *adaptive control algorithm* and describes the (σ, ρ, λ) regulator. Section 4 presents the theorems for the *WDB*, the *input rate threshold*, and the *worst-case delay improvement* of the single regulated end host. The theoretical analysis for EMcast is presented in Section 5. Section 6 uses the simulations to observe the WDB performance for a regulated end host and for different EMcast schemes. Section 7 concludes the paper.

2 Related Work

Traffic control has been studied for the applications with various constraints in speed, quality, and consistency of data delivery. Not many mature researches have been done to address the multicast traffic control and most of these researches are designed for the IP multicast.

2.1 Traffic Control in IP Multicast

IP multicast usually employs open-loop or feedback traffic control systems. In an open-loop control system [31], a predetermined control strategy that a session makes resource reservations ahead of time is fixed. The senders control their sending rate within the reservation and do not response to changing network conditions. The open-loop control is difficult to implement because the Internet provides best effort services without service reservation. Most multicast traffic control schemes (for example, Representative [17], Random Listening Algorithm (RLA) [18], Multicast Transmission Control Protocol (MTCP) [20], and Golestani and Sabnani [21]) are based on feedback control. In a feedback control system, the control parameter is adjusted on the fly. The control result reflecting the instantaneous network situations is measured and sent back to the associate node (for example, the sender), who will then adjust the transmission accordingly. TCP-Friendly Rate Control (TFRC) [36] is a feedback control mechanism designed to compete with TCP traffic for bandwidth in unicast Internet environment. The TFRC receiver calculates the congestion control information (that is, the loss rate) and feedbacks the information to the sender, who then measures the round-trip time (RTT) and gives the acceptable transmit rate. Such receiver-based mechanism enables TFRC to be more easily extended to a multicast traffic control (TCP-Friendly Multicast Congestion Control (TFMCC) [19]). In TFMCC, each receiver measures the loss rate and its RTT to the sender and then decides a sending rate based on the equation for the TCP throughput. The sender selects the receiver who reports the lowest rate as the current limiting receiver.

IP multicast distributes packets to the group's multicast address instead of each group member's individual IP address. During network communications, different hosts have different instantaneous capacities. The traffic control in the IP multicast is usually complex in order to guarantee the communication qualities of end hosts with different capacities. For example, TFMCC excessively employs the benefit of the TCP layer. Furthermore, to detect the current network state, considerable feedback overheads are introduced into the network.

2.2 Traffic Control in EMcast

Several end-to-end TCP-friendly multicast traffic control algorithms [28], [29] have been studied for EMcast. Amir and Danilov [31] proposed OverlayTFMRC focusing on quality-of-service (QoS) transport support for multimedia streaming and scalable TFMCC. Although the *clustered-receiver-based* random delay strategy is employed, the scheme cannot remove the control overheads yet. Actually, many large-scale practical systems (for example, Overcast [10] and Application Level Multicast Infrastructure (ALMI) [4]) implement congestion control in the overlay path between each pair of nodes. These systems are actually implicitly TCP friendly, where the overlay paths are constructed through using TCP. Generally, because the sending sources employ the lowest sending rate calculated by the receivers [30], the throughput in the system with the TCP-friendly control scheme decreases with the increasing number of receivers. Therefore, in [31], the hop-by-hop control, rather than the end-to-end technique, is suggested to be implemented because of the end hosts' storing and forwarding functions in EMcast.

Keller and Biersack [32] proposed a fixed-size window-based traffic control protocol for adjacent nodes. Each end host maintains a buffer for each of its outgoing interface. A fixed window control is applied to prevent the host from forwarding any packets to another host with the full-sized window. Other typical EMcast traffic control schemes include Cost Benefit [33] and ROMA [34]. Basically, these schemes implement the quasi hop-by-hop control and lack of the efficiency to support real-time streaming media. A single slow receiver can degrade the performance of the entire system. When it comes to the multiple group environment, the situation becomes worse because there is no coupled process for different group streams.

Our proposed algorithm implements traffic control on a hop-by-hop basis. Without introducing feedback overheads, each host adaptively decides its control models based on the traffic input rates. Furthermore, each receiver implements the traffic control based on its own capacity and, therefore, the slow receiver problem is solved.

3 Adaptive Control Algorithm

In this section, we introduce how the *adaptive control algorithm* works for traffic control in overlay networks.

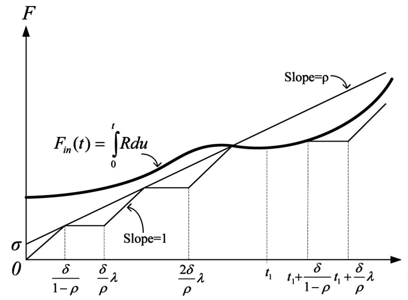


Fig. 2. An example of the operations of a (σ, ρ, λ) regulator. The x-axis represents the communication time, and the y-axis represents the amount of data from the stream since the beginning of the communication. R is the traffic instantaneous input rate and $\lambda = \frac{1}{(1-\rho)}$ is the traffic control factor of the (σ, ρ, λ) regulator.

The model of the novel (σ, ρ, λ) regulator is presented. Also, the values of some parameters (for example, λ and *regulator period*) are defined and calculated.

We first introduce the communication environment of the algorithm. Similar to [15] and [16], end hosts in the multigroup network are equipped with multiplexers (muxes) to control the input flows. The function of each mux is to merge the flows arriving at its two or more input links into its single output link. In our traffic service algorithm, the general mux is considered. A general mux is a mux such that a packet of one flow may have priority over a packet of another flow for transmission. For brevity, we define an end host consisting of a mux that is regulated by a $(\sigma, \rho, \lambda)/(\sigma, \rho)$ regulator on each of its input links as a $(\sigma, \rho, \lambda)/(\sigma, \rho)$-regulated end host. Suppose there are K groups with $n_I (I \in [1, K])$ end hosts each. Denote each group as G^I $(I \in [1, K])$. Without loss of generality, a member g_j^i in the multigroup network may join in $\hat{K}$ groups ($\hat{K} \in [1, K]$, $i \in [1, \hat{K}]$, $j \in [1, n_i]$, and n_i is the size of the ith group that is denoted as G^i). Assume that there is only one real-time flow in each group. We denote the ith flow's burst data amount as σ_i and the long-term average input rate as ρ_i. Therefore, there is a total of K flows in the multigroup network. For simplicity, in this paper, we assume that each link in the network has a uniform available capacity $C = 1$. When the assumption is released, the theorems and their proofs can be similarly developed by multiplying σ_i and ρ_i by C. The inequality $\sum_{i=1}^{\hat{K}} \rho_i \leq 1$ at each end host g_j^i is regarded as the *stability condition* of the multigroup network. For $\hat{K}$ homogeneous flows with the input rate bound ρ, the *stability condition* at each group member can be simplified as $\hat{K}\rho \leq 1$.

The basic idea of the *adaptive control algorithm* is that each end host adaptively employs the same traffic control model as the (σ, ρ) regulator under the normal traffic load situation, but it provides more regulations by using new (σ, ρ, λ) regulators in the overlay network to control the traffic output under the heavy traffic load situation. Fig. 2 gives the operations of the (σ, ρ, λ) regulator serving for one of the K flows. As illustrated by the zigzag curve, the (σ, ρ, λ) regulator blocks the flow's output for V time units after outputting the flow for W time units. We call the V time units (that is, the horizontal parts of the zigzag curve) as the flow's vacation period, and the W time units (that is, the sloping parts of the zigzag curve) as the flow's working period. Other flows will experience the similar operations through their (σ, ρ, λ) regulators. In order to smooth the simultaneous burstiness of $\hat{K}$ flows, the *adaptive control algorithm* at each end host enables one regulator to work for its flow at each time in turn, whereas other regulators block their flows at the same time. The period of $(V + W)$ time units is defined as one *regulator period* of the flow and is equal to $\frac{\sigma}{\rho}\lambda$. We will explain the physical rationalness of $\frac{\sigma}{\rho}\lambda$ later in this section. λ is a new parameter employed by our regulator. It decides the flow's vacation period and working period. We now see how we can decide W_i, V_i, and λ_i for the flow from the ith group.

Denote the $(\sigma, \rho)/(\sigma, \rho, \lambda)$ regulator of the ith flow at g_j^i as the $(\sigma_i, \rho_i)/(\sigma_i, \rho_i, \lambda_i)$ regulator, respectively, and the instantaneous input rate of the ith flow at g_j^i as R_i. Similar to the (σ_i, ρ_i) regulator in [15] and [16], our $(\sigma_i, \rho_i, \lambda_i)$ regulator considers that $R_i \sim (\sigma_i, \rho_i)$. The term $R_i \sim (\sigma_i, \rho_i)$ holds when $\int_{t_1}^{t_2} R_i dt \leq \sigma_i + \rho_i(t_2 - t_1)$ satisfies, where σ_i and ρ_i are the ith flow's burst data amount and long term average input rate, respectively, and t_1 (t_2) is the communication time, with $t_2 \geq t_1$. $R_i \sim (\sigma_i, \rho_i)$ shows the physical meaning that the input amount of the ith flow in any interval is upper bounded by the burst data amount plus the product of the long-term average input rate and the length of the interval.

In order to guarantee that the total amount of traffic output at the end host g_j^i should not be greater than the total number of the input traffic in the regulator during the period of m of W_i and $(m-1)$ of V_i, g_j^i's output should satisfy $mW_i \leq \sigma_i + [mW_i + (m-1)V_i]\rho_i$. In Fig. 2, the cross points of the zigzag curve and the trend line indicate the time that all of the blocked data from the flow are output by the (σ, ρ, λ) regulator. Furthermore, because all of the output capacity $C = 1$ is occupied by the flow, the value of the slope of the (σ, ρ, λ) regulator curve is 1. Based on this analysis, we achieve that $W_i = \frac{\sigma_i}{(1-\rho_i)}$. It infers that $\frac{m\sigma_i}{1-\rho_i} \leq \sigma_i + [\frac{(m-1)\lambda_i\sigma_i}{\rho_i} + \frac{\sigma_i}{1-\rho_i}]\rho_i$. That is, $\lambda_i \geq \frac{1}{1-\rho_i}$. Since $V_i = \frac{\sigma_i}{\rho_i}\lambda_i - W_i$, a smaller λ_i generates a shorter vacation period. Therefore, considering the reduction of the worst-case delay, we have

$$\lambda_i = \frac{1}{1 - \rho_i}. \quad (1)$$

Equation (1) infers $V_i = \frac{\sigma_i}{\rho_i}$.

We now analyze the physical meaning of *regulator period*. For brevity, we suppose there are $\hat{K}$ homogeneous flows (that is, $\rho_i = \rho$, $i \in [1, \hat{K}]$, and $j \in [1, n_i]$). By the stability condition, we assume that $\rho \to \frac{1}{\hat{K}}$ in the worst case. Then, we have $V = \frac{\sigma}{\rho} \approx \hat{K}\sigma = \frac{(\hat{K}-1)\sigma}{(1-\frac{1}{\hat{K}})} \approx (\hat{K} - 1)W$. It implies that, when the input rate on each link is very high, the vacation interval of each regulator is nearly the same as the summation of the working intervals of other $(\hat{K} - 1)$ regulators. Therefore, the introduction of the *regulator period* and vacation has the physical rationalness. The detailed operations of the *adaptive control algorithm* are given as follows:

Adaptive Control Algorithm

Input: the input rate threshold ρ_j^* of the member g_j^i who joins in $\hat{K}$ groups $G^i = \{g_1^i, \ldots, g_j^i, \ldots, g_{n_i}^i\}$;
// $i \in [1, \hat{K}]$, $j \in [1, n_i]$, n_i is the size of G^i
Output: traffic control model;
1. End host g_j^i calculates the average input rate $\bar{\rho}$ of $\hat{K}$ real-time flows that belong to $\hat{K}$ groups, respectively;
2. If $(\bar{\rho}_j \in (0, \rho_j^*))$ {
 g_j^i employs the same traffic control model as the (σ_i, ρ_i) regulator; }
3. Else if $(\bar{\rho}_j \in [\rho_j^*, \frac{1}{\hat{K}}))$ {
 g_j^i employs the $(\sigma_i, \rho_i, \lambda_i)$ regulators to control the output of $\hat{K}$ input flows by the following steps alternatively:
 (1) *On state*: it works in a work-conserving way for $W_i = \frac{\sigma_i}{1-\rho_i}$ time units and then
 (2) *Off state*: it takes a vacation of $V_i = \frac{\lambda_i\sigma_i}{\rho_i} - \frac{\lambda_i}{1-\rho_i}$ time units by turning off the input of the ith flow at the end host g_j^i.}

The selection of the traffic control model is based on the flows' instantaneous input rates. The input rates indicate the instantaneous situations of underlying links. More specifically, the input rate $\rho_i (i \in [1, \hat{K}])$ at an end host indicates the minimum instantaneous capacity among all underlying links that are covered by the overlay paths connecting the flow sender and the end host. The calculations of W_i and V_i are based on ρ_i and, therefore, based on the instantaneous capacity of underlying network links. The algorithm "intelligently" judges whether the end host is in the face of congestion or not according to the average rate of all input flows. When the average rate is larger than the rate threshold ρ^* (that is, the end host has no enough output capacity to work for all received flows at the same time), the algorithm "intelligently" blocks the simultaneous entering flows at a short specific period in turn. It can be seen that the key problem of the *adaptive control algorithm* is to find the input rate threshold ρ^* at which the algorithm should change the traffic control model. We will prove the existence and address the calculation of ρ^* through the theoretical analysis later.

4 Analysis of Worst-Case Delay Bound for the Single Regulated End Host

We analyze the WDB for the single regulated end host in this section. The results obtained will serve as an important basis of WDB analysis as the packets pass through the EMcast tree.

The following lemma characterizes the delay of any input flow with the rate function of $R \sim (\sigma^*, \rho)$ at the (σ, ρ, λ)-regulated end host.

Lemma 1. *If the rate function R of the input flow satisfies the burst constraint of the (σ^*, ρ) regulator, that is, $R \sim (\sigma^*, \rho)$, then the delay incurred by the (σ, ρ, λ) regulator is upper bounded by*

$$D = \frac{(\sigma^* - \sigma)^+}{\rho} + \frac{2\lambda\sigma}{\rho}. \quad (2)$$

Proof. To prove the lemma, it is assumed that there exists $\tilde{R}_0$ that satisfies the traffic constraint of the (σ, ρ) regulator, that is, $\tilde{R}_0 \sim (\sigma, \rho)$. We now consider two cases.

In case $\sigma^* \leq \sigma$, obviously, the largest backlog occurs at each end of a vacation. Without loss of generality, let $B(s)$ (s is an integer) denote the backlog of the regulator at time $\frac{s\lambda\sigma}{\rho}$, which is the end of a vacation. By the burst constraint of R, there is $B(0) \leq \sigma$. We may infer that $B(s) \leq (1+\lambda)\sigma$ for all $s \geq 0$ by the following induction on s. For simplicity, we denote the input flow rate bound without burstiness as ρ. At time $\frac{s\lambda\sigma}{\rho}$, the traffic arriving at the (σ, ρ, λ) regulator is $\lambda\sigma$ during the period of $\frac{\lambda\sigma}{\rho}$. On the other hand, since $\sigma^* \leq \sigma$, at time $\frac{s\sigma}{1-\rho}$, the regulator can output the amount of traffic that equals the amount of input traffic during the period $[\frac{(s-1)\sigma}{1-\rho}, \frac{(s-1)\sigma}{\rho}\lambda]$. That is to say, from the beginning of the communication to the time $\frac{s\lambda\sigma}{\rho}$, the maximum total backlog is $\lambda\sigma$ (that is, the amount of traffic entered during the period $[\frac{s\sigma}{1-\rho}, \frac{s\lambda\sigma}{\rho}]$). Based on this and considering the induction assumption $B(0) \leq \sigma$, we can infer that $B(s) \leq (1+\lambda)\sigma < 2\lambda\sigma$. Because $B(s)$ may be output by the regulator at the rate of ρ, the maximum delay could be as long as $\frac{2\lambda\sigma}{\rho}$.

In case $\sigma^* > \sigma$, because $\tilde{R}_0 \sim (\sigma, \rho)$, it can be seen that the regulator may take some additional time to process the burst traffic $(\sigma^* - \sigma)$ originating from the input flow with the rate ρ. Therefore, the delay is $(\sigma^* - \sigma)/\rho$. Taking the two cases into consideration, we have the delay bound for the (σ, ρ, λ) regulator $D = \frac{(\sigma^*-\sigma)^+}{\rho} + \frac{2\lambda\sigma}{\rho}$. □

4.1 Worst-Case Delay Bound

In this section, we present two theorems for the WDBs with K heterogeneous (Theorem 1) and homogeneous (Theorem 2) real-time flows, respectively, by applying Lemma 1 in the $(\sigma_i, \rho_i, \lambda_i)$-regulated general mux, with $1 \leq i \leq K$.

Theorem 1. *Let the rate function of the input flow f_i be given by R_i such that $R_i \sim (\sigma_i, \rho_i)$, $1 \leq i \leq K$, and $\sigma_i^* = \rho_i(1-\rho_i)$. $\min_{1\leq j\leq K}\{\frac{\sigma_j}{\rho_j(1-\rho_j)}\}$. Then, the maximum delay experienced by a traffic bit in a general mux with the $(\sigma_i^*, \rho_i, \lambda_i)$ regulator is upper bounded by*

$$\hat{D}_g = \sum_{i=1}^{K} \frac{\sigma_i^*}{1-\rho_i} + 2\min_{1\leq i\leq K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} + \max_{1\leq i\leq K}\left\{\frac{\sigma_i - \sigma_i^*}{\rho_i}\right\}.$$

Proof. Without loss of generality, the delay experienced by any traffic bit from the flow $f_j (j \in [1, K])$ is upper bounded by $\hat{D}_g \leq D_1 + D_2$, where D_1 is the delay experienced by the bit passing through the corresponding regulator, and D_2 is the delay bound of the mux. By Lemma 1 and $\lambda_i = \frac{1}{1-\rho_i}$, there exists

$$\begin{aligned} D_1 &\leq \frac{2\lambda_i\sigma_i^*}{\rho_i} + \max_{1\leq i\leq K}\left\{\frac{\sigma_i - \sigma_i^*}{\rho_i}\right\} \\ &= 2\min_{1\leq i\leq K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} + \max_{1\leq i\leq K}\left\{\frac{\sigma_i - \sigma_i^*}{\rho_i}\right\}. \end{aligned}$$

It can be seen that the amount of data bits from any flow f_i arriving at the mux in any period of $\min_{1\leq i\leq K}\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\}$ time units is upper bounded by $P^{(i)} = \frac{\sigma_i^*}{1-\rho_i}$; hence, the total amount of data bits arriving at the mux in any period of $\min_{1\leq i\leq K}\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\}$ time units is not more than $\sum_{i=1}^{K} P^{(i)} = \sum_{i=1}^{K} \frac{\sigma_i^*}{1-\rho_i}$.

Since the mux is work conserving, with service rate $C = 1$, the above inequality means that each backlog at the mux at any time is upper bounded by $D_2 = \sum_{i=1} K \frac{\sigma_i^*}{1-\rho_i}$. In other words, it is the upper bound on delay for any bit passing through the mux. Thus, the theorem is proved. □

Theorem 2 gives the WDBs of K homogeneous real-time flows passing through the (σ, ρ, λ)-regulated general mux.

Theorem 2. *For a regulated general mux with K homogeneous input flows, let the input traffic rate functions be R_i such that $R_i \sim (\sigma_0, \rho)$, $1 \leq i \leq K$, and $\rho \leq \frac{1}{K}$. Then, the maximum delay experienced by any data bit in a (σ, ρ, λ)-regulated general mux is upper bounded by*

$$\hat{D}_g = \frac{K\sigma}{1-\rho} + \frac{(\sigma_0 - \sigma)^+}{\rho} + \frac{2\lambda\sigma}{\rho}. \tag{3}$$

The proof of Theorem 2 is similar to that of Theorem 1 and, thus, is omitted here.

Remark 1. Based on [15, (13)], if a general mux has K heterogeneous (homogeneous) input flows, and the rate function for each flow is given by $R_i \sim (\sigma_i, \rho_i)$ and $\sum_{1\leq i\leq K} \rho_i \leq 1$ (R_i such that $R_i \sim (\sigma_0, \rho)$ and $\rho \leq \frac{1}{K}$), then the maximum delay in the general mux is upper bounded by

$$D_g = \frac{\sum_{1\leq i\leq K}\sigma_i}{1 - \sum_{1\leq i\leq K}\rho_i}\left(D_g = \frac{K\sigma_0}{1-K\rho}\right). \tag{4}$$

4.2 Input Rate Threshold ρ^*

Now, we are going to derive the control threshold ρ^* for our *adaptive control algorithm* to distinguish the high-rate real-time traffic from the normal rate traffic. We give the following notations:

$$\begin{aligned} &\xi_{max} = \max_{1\leq i\leq K}\{\rho_i(1-\rho_i)\}, \xi_{min} = \min_{1\leq i\leq K}\{\rho_i(1-\rho_i)\}, \\ &\rho_{min} = \min_{1\leq i\leq K}\{\rho_i\}, \bar{\rho} = \left(\sum_{i=1}^{K}\rho_i\right)\Big/ K. \end{aligned} \tag{5}$$

We then introduce a condition that will be employed by the following inference:

$$\frac{\xi_{max} - \xi_{min}}{\xi_{max}} \leq \frac{\rho_{min}}{\bar{\rho}}. \tag{6}$$

Theorem 3. *Assume that a $(\sigma_i^*, \rho_i, \lambda_i)$-regulated mux with the general service discipline has K input links, with the rate function for each link given by R_i, such that $R_i \sim (\sigma_i, \rho_i)$, $1 \leq i \leq K$, and $\sum_{i=1}^{K}\rho_i \leq 1$. If $K \geq 2$ and (6) is satisfied, then there exists a rate threshold $0 < \rho^* < \frac{1}{K}$ such that*

1. *if $\rho^* \leq \bar{\rho} < \frac{1}{K}$, $\hat{D}_g \leq D_g$, and if $0 < \bar{\rho} \leq \rho^*$, $D_g \leq \hat{D}_g$, where $\hat{D}_g$ and D_g are the WDBs of the real-time flows constrained by the $(\sigma_i^*, \rho_i, \lambda_i)$-regulated general mux and the (σ_i^*, ρ_i)-regulated general mux, respectively, and $\bar{\rho}$ is the average input rate of K flows and*

2. *when K is large enough, the ratio of the range (called the control range) $[\rho^*, \frac{1}{K})$ to the total range $(0, \frac{1}{K})$ is approximately given by $\frac{1/K-\rho^*}{1/K} \approx \frac{5-\sqrt{21}}{1} \approx 0.21$.*

Proof. 1) By condition (6), for each part of the expression of D_g in Theorem 1, assuming that $\sigma = \min_{1\le i\le K}\{\sigma_i\}$, we have

$$\sum_{i=1}^{K} \frac{\sigma_i^*}{1-\rho_i} = \sum_{i=1}^{K} \rho_i \min_{1\le j\le K}\left\{\frac{\sigma_j}{\rho_j(1-\rho_j)}\right\} \le \frac{(\sum_{i=1}^{K}\rho_i)\sigma}{\xi_{max}},$$

$$\min_{1\le i\le K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} \le 2\frac{\sigma}{\xi_{max}}, \max_{1\le i\le K}\left\{\frac{\sigma_i-\sigma_i^*}{\rho_i}\right\}$$

$$= \max_{1\le i\le K}\left\{\frac{\sigma_i - \sigma\rho_i(1-\rho_i)\frac{1}{\xi_{max}}}{\rho_i}\right\}$$

$$\le \frac{\sigma[\frac{\sigma_i}{\sigma} - \frac{\xi_{min}}{\xi_{max}}]}{\rho_{min}} = \frac{\xi_{max}-\xi_{min}}{\xi_{max}}\frac{\sigma}{\rho_{min}} + \frac{\sigma_i - \sigma}{\rho_{min}}.$$

Then, D_g in Theorem 1 can be rewritten as

$$\hat{D}_g \le \frac{\left(\sum_{i=1}^{K}\rho_i\right)\sigma}{\xi_{max}} + \frac{2\sigma}{\xi_{max}} + \frac{\xi_{max}-\xi_{min}}{\xi_{max}}\cdot\frac{\sigma}{\rho_{min}} + \frac{\sigma_i-\sigma}{\rho_{min}}. \quad (7)$$

Note that $h(x) = x(1-x)$ is an increasing function in the interval $[0, 1/K)$ when $K \ge 2$; thus, for $\rho_i \in [0, 1/K)$, we have $\xi_{max} = \max_{1\le i\le K}\{\rho_i(1-\rho_i)\} \ge \bar{\rho}(1-\bar{\rho})$.

With (5), (6), and (7), we have[1]

$$\hat{D}_g = \frac{K\sigma}{1-\bar{\rho}} + \frac{2\sigma}{\bar{\rho}(1-\bar{\rho})} + \frac{\sigma}{\bar{\rho}} + \frac{1}{\rho_{min}}. \quad (8)$$

On the other hand, D_g in (3) can be represented as $D_g = \frac{K\sigma}{1-K\bar{\rho}}$.

Let

$$g_1(\bar{\rho}) = \frac{K}{1-\bar{\rho}} + \frac{2}{\bar{\rho}(1-\bar{\rho})} + \frac{1}{\bar{\rho}}, g_2(\bar{\rho}) = \frac{K}{1-K\bar{\rho}}. \quad (9)$$

Considering the equation

$$g_1'(\bar{\rho}) = \frac{K}{(1-\bar{\rho})^2} - \frac{2(1-2\bar{\rho})}{\bar{\rho}^2(1-\bar{\rho})^2} - \frac{1}{\bar{\rho}^2} = 0,$$

with the positive solution by $\bar{\rho}_0 = \frac{-3+\sqrt{9+3(K-1)}}{K-1}$, it is clear that $\bar{\rho}_0$ is the minimum point of the function $g_1(\bar{\rho})$. Thus, the function $g_1(\bar{\rho})$ increases in $[\bar{\rho}_0, 1)$ such that $lim_{\bar{\rho}\to 1} g_1(\bar{\rho}) = +\infty$ and it decreases in $(0, \bar{\rho}_0]$ such that $\lim_{\bar{\rho}\to 0} g_1(\bar{\rho}) = +\infty$. Since $g_2'(\bar{\rho}) \ge g_1'(\bar{\rho})$, $0 < \bar{\rho} < \frac{1}{K}$, it can be inferred that $g_1(\bar{\rho}) = g_2(\bar{\rho})$ has a unique positive solution ρ^* such that $0 < \rho^* < 1/K$. Consequently, $g_1(\bar{\rho}) \le g_2(\bar{\rho})$ when $\bar{\rho} \in [\rho^*, 1/K)$, and $g_1(\bar{\rho}) \ge g_2(\bar{\rho})$ when $\bar{\rho} \in (0, \rho^*]$. Thus, 1) is proved.

2) By 1), ρ^* is the unique positive solution of $g_1(\bar{\rho}) = g_2(\bar{\rho})$, which can be deduced to

$$(K^2 - 2K)\bar{\rho}^2 + (3K+1)\bar{\rho} - 3 = 0.$$

By solving this equation, we have

$$\rho^* = \frac{-(3K+1) + \sqrt{(3K+1)^2 + 12(K^2-2K)}}{2(K^2-2K)}.$$

It is easy to see that $\lim_{K\to\infty}\frac{1/K-\rho^*}{1/K} = \lim_{K\to\infty}(1-K\rho^*) = \frac{5-\sqrt{21}}{2}$. Since it has been assumed that $C = 1$, 2) thus holds. □

Theorem 4 gives the rate threshold ρ^* for the single regulated end host with K homogeneous flows.

Theorem 4. *Assume that a (σ, ρ, λ)-regulated mux with the general service discipline has K input links, with the rate function for each link given by R_i, such that $R_i \sim (\sigma_0, \rho)$, $1 \le i \le K$, and $\rho \le 1/K$. When $K \ge 2$, there exists a rate threshold $0 < \rho^* < 1/K$ such that*

1. *if $\rho^* \le \rho < \frac{1}{K}$, $\hat{D}_g \le D_g$, and if $0 < \rho \le \rho^*$, $D_g \le \hat{D}_g$, where $\hat{D}_g$ and D_g are the WDBs of the real-time flows constrained by the (σ, ρ, λ)-regulated general mux and the (σ, ρ)-regulated general mux, respectively, and*
2. *when K is large enough, the ratio of the range $[\rho^*, 1/K)$ with respect to the overall range $(0, 1/K)$ is about $\frac{\frac{1}{K}-\rho^*}{\frac{1}{K}} \approx 2 - \sqrt{3} \approx 0.27$.*

The proof of Theorem 4 can be similarly established as the proof of Theorem 3 and, thus, is omitted here.

4.3 Improvement of Worst-Case Delay Bound

We now analyze the WDB improvement of the (σ, ρ, λ) regulator over the (σ, ρ) regulator for the heterogeneous (Theorem 5) and the homogeneous (Theorem 6) real-time flows, respectively. As we will see, the worst-case delay in the $(\sigma_i, \rho_i, \lambda_i)$-regulated general mux, with $1 \le i \le K$, can be reduced effectively when the average rate $\bar{\rho}$ of K input flows is above the input rate threshold ρ^*.

Theorem 5. *Let the rate functions of the input traffic be given by $R_i \sim (\sigma, \rho_i)(1 \le i \le K)$, with $\sum_{1\le i\le K}\rho_i \le 1$, and D_g and $\hat{D}_g$ be the WDBs for a general mux regulated by the (σ_i, ρ_i) and $(\sigma_i, \rho_i, \lambda_i)$ regulators, respectively. When the number of input links $K \ge 2$, for any positive integer n such that $\frac{1}{K} - \frac{1}{K^{n+1}} \ge \rho^*$, we have $\frac{D_g}{\hat{D}_g} \ge O(K^n)$ whenever $\bar{\rho} \in [\frac{1}{K} - \frac{1}{K^{(n+1)}}, \frac{1}{K})$.*

Proof. By Theorem 1 and Remark 1, when the general mux is regulated by the (σ_i, ρ_i) and $(\sigma_i, \rho_i, \lambda_i)$ regulators, the WDBs are expressed as

$$D_g = \frac{\sum_{1\le i\le K}\sigma_i}{1 - \sum_{1\le i\le K}\rho_i}$$

and

$$\hat{D}_g = \sum_{i=1}^{K}\frac{\sigma_i^*}{1-\rho_i} + 2\min_{1\le i\le K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} + \max_{1\le i\le K}\left\{\frac{\sigma_i-\sigma_i^*}{\rho_i}\right\},$$

respectively.

When K is large enough, by Theorem 3, we can prove that $\rho^* \approx \frac{\sqrt{21}-3}{2K}$. Thus, when n is chosen properly, the inequality $\frac{1}{K} - \frac{1}{K^{n+1}} \ge \rho^*$ holds. Then, for

1. By (7), with $\xi_{max} \ge \bar{\rho}(1-\bar{\rho})$ and $K\bar{\rho} = \sum_{i=1}^{K}\rho_i$, for the first part in (7), we have $\frac{(\sum_{i=1}^{K}\rho_i)\sigma}{\xi_{max}} \le \frac{K\bar{\rho}\sigma}{\bar{\rho}(1-\bar{\rho})} = \frac{K\sigma}{1-\bar{\rho}}$. With $\xi_{max} \ge \bar{\rho}(1-\bar{\rho})$, for the second part in (7), we have $\frac{2\sigma}{\xi_{max}} \le \frac{2\sigma}{\bar{\rho}(1-\bar{\rho})}$. With $\frac{\xi_{max}-\xi_{min}}{\xi_{max}} \le \frac{\rho_{min}}{\bar{\rho}}$, for the third part in (7), we have $\frac{\xi_{max}-\xi_{min}}{\xi_{max}}\cdot\frac{\sigma}{\rho_{min}} \le \frac{\sigma}{\bar{\rho}}$. With $\sigma_i \le 1$, $\sigma \le 1$, and $\sigma_i \ge \sigma$, for the fourth part in (7), we have $\frac{\sigma_i-\sigma}{\rho_{min}} \le \frac{1}{\rho_{min}}$. Therefore, in the worst case, it can be inferred that $\hat{D}_g = \frac{K\sigma}{1-\bar{\rho}} + \frac{2\sigma}{\bar{\rho}(1-\bar{\rho})} + \frac{\sigma}{\bar{\rho}} + \frac{1}{\rho_{min}}$.

any $\bar{\rho} \in [\frac{1}{K} - \frac{1}{K^{n+1}}, \frac{1}{K})$, it is easy to infer that $\bar{\rho} \in [\rho^*, \frac{1}{K})$, and we have

$$\frac{D_g}{\hat{D}_g} \geq \frac{K\bar{\rho}(1-\bar{\rho})}{(1-K\bar{\rho})[3+(K-1)\bar{\rho}]} \geq \frac{(1-\frac{1}{K^n})(1-\frac{1}{K})K^n}{4} = O(K^n).$$

□

For K homogeneous flows, we give the worst-case delay improvement in Theorem 6.

Theorem 6. *Let the input rate function R_i of the homogeneous flows be the same as the above theorem and let D_g and $\hat{D}_g$ be the WDBs for a general mux regulated by the (σ, ρ) and the (σ, ρ, λ) regulators, respectively. When the number of input links $K \geq 2$, there exists a rate threshold $0 < \rho^* < 1/K$ for any n such that $\frac{1}{K} - \frac{1}{K^{n+1}} \geq \rho^*$. We have $\frac{D_g}{\hat{D}_g} = O(K^n)$, whenever $\rho \in [\frac{1}{K} - \frac{1}{K^{n+1}}, \frac{1}{K})$.*

Theorem 6 can be proved in the similar way as we prove Theorem 5 and, thus, we omit it here.

5 Analysis of Worst-Case Delay Bound for the End Host Multicast

Based on the above theorems, we achieve the theoretical results on the *WDB*, the *input rate threshold*, and the *WDB improvement* for the regulated EMcast tree in this section. In our analysis, we use the DSCT tree, as in [18], as the model of EMcast. DSCT arranges the end hosts in each group $G^I (I \in [1, K])$ to construct a DSCT multicast tree. The DSCT tree is a location-aware hierarchy and cluster tree architecture. It partitions the group members into different *local domains*. Each *local domain* only contains the group members attaching to the same backbone routers. In terms of the *RTT value*, the closest s_{ina} group end hosts are assigned into the same "intracluster." As expressed in [14, (1)], the "intracluster" size s_{ina} is a random integer between k and $3k-1$ if the number of unassigned members is greater than $3k-1$; otherwise, s_{ina} is the number of the *unassigned* group members. Each cluster has a cluster core that joins in the immediate upper layer and forms clusters in this layer with other $(s_{ina}-1)$ closest cluster cores. Each *local domain* has a *local core* who is the end host in the upmost layer of the *local domain*. For the connections of different *local domains*, the closest *local cores* form "interclusters" with the size s_{ine} that is a random integer between k and $3k-1$, as expressed in [14, (2)]. The local cores then continue constructing upper layers by the same way to layer the end hosts in each *local domain*. We first analyze the height bound H of the DSCT tree in Lemma 2 when there are n members in the group.

Lemma 2. *For a multicast group with n members, the height of the DSCT tree constructed by the n members is upper bounded by*

$$H = \left\lceil \log_k^{[k+(n-j_1)(k-1)]} \right\rceil, \tag{10}$$

where k is a random integer that is decided by the group size and the application requirements (k is set as 3 in the computer experiments [8]), and $j_1 (0 \leq j_1 \leq k-1)$ is the number of the last unassigned members in the lowest layer L_1 of the DSCT tree.

Proof. According to [14, (1) and (2)], the n members will construct the highest DSCT tree when the sizes of all clusters equal k.

Suppose the DSCT tree has l layers. We use i_1 to denote the number of clusters with the size k in the lowest layer L_1 and use j_1 to denote the remaining members who have not joined in any of the i_1 clusters. It can be inferred that $i_1 = \lfloor \frac{n}{k} \rfloor$, and $0 \leq j_1 \leq k-1$. The j_1 members will form a new cluster in L_1. Hence, there are at most (i_1+1) clusters in L_1. We have $n = i_1 k + j_1$.

Because the core of each cluster joins in the immediate upper layer L_2, we can infer $i_1 + 1 = i_2 k + j_2$, where $i_2 = \lfloor \frac{i_1+1}{k} \rfloor$ is the number of clusters with the size k in L_2, and $j_2 \in [0, k-1]$ is the number of members who have not joined in any of the i_2 clusters. Similarly, in the layer L_l, we can derive

$$i_{l-1} + 1 = i_l k + j_l, \tag{11}$$

where $i_l = \lfloor \frac{i_{l-1}+1}{k} \rfloor$ is the number of clusters with the size k, and $j_l \in [0, k-1]$ is the number of members who have not joined in the i_l clusters.

Based on the above equations, using the iteration, we have

$$\begin{aligned} n &= j_1 + i_1 k = j_1 + (j_2 - 1 + i_2 k)k = j_1 + (j_2-1)k + i_2 k^2 \\ &= \ldots = j_1 + (j_2-1)k + (j_3-1)k^2 + \ldots + (j_l-1)k^l + i_l k^{l+1}. \end{aligned} \tag{12}$$

Because there is only one member in the highest layer L_l, we have $i_l = 0$ and $j_l = 1$. Furthermore, (12) shows that the tree will have the maximum layer number when $j_2 = j_3 = \ldots = j_{l-1} = 2$. Thus, we can infer from (12) that

$$n = j_1 + k + k^2 + \ldots + k^{l-1} = j_1 + \frac{k - k^l}{1-k}. \tag{13}$$

It can be achieved from (13) that $l = \lceil \log_k^{[k+(n-j_1)(k-1)]} \rceil$. In other words, the height of the DSCT tree that covers n members is upper bounded by $H = \lceil \log_k^{[k+(n-j_1)(k-1)]} \rceil$.□

By applying Lemma 2, we analyze the WDBs of EMcast with K heterogeneous flows and K homogeneous flows in Theorem 7 and Theorem 8, respectively.

Theorem 7. *Suppose there are K groups in the regulated multigroup network, and each group has $n_i (i \in [1, K])$ end hosts that construct a DSCT tree. If one group has one real-time flow, and the flow is constrained by the rate function R_i such that $R_i \sim (\sigma_i, \rho_i)$, with the stability condition $\sum_{i=1}^{\hat{K}} \rho_i \leq 1$ at each end host who joins in $\hat{K} (\hat{K} \in [1, K])$ groups, let $\lambda_i = \frac{1}{1-\rho_i}$, and $\sigma_i^* = \rho_i (1-\rho_i) \min_{1 \leq j \leq \hat{K}} \{\frac{\sigma_j}{\rho_j (1-\rho_j)}\}$, then*

1. *The maximum multicast delays experienced by any bit passing through the multigroup network with the $(\sigma_i^*, \rho_i, \lambda_i)$-regulated general MUXes are upper bounded by*

$$\hat{D_{mg}} = \sum_{i=1}^{K} \frac{(\hat{H}-1)\sigma_i^*}{1-\rho_i} + 2 \min_{1 \leq i \leq K} \left\{ \frac{(\hat{H}-1)\sigma_i}{\rho_i (1-\rho_i)} \right\} + \max_{1 \leq i \leq K} \left\{ \frac{(\hat{H}-1)(\sigma_i - \sigma_i^*)}{\rho_i} \right\},$$

where $\hat{H} = \max_{1 \le i \le K}\{H_i\}$, *and* H_i *is the height bound of the DSCT tree in the group* G^i *that can be derived by Lemma 2.*

2. *If* $K \ge 2$ *and (6) is satisfied, then there exists a rate threshold* $0 < \rho^* < \frac{1}{K}$ *such that* $\hat{D_{mg}} \le D_{mg}$ *if* $\rho^* \le \rho < \frac{1}{K}$, *and* $D_{mg} \le \hat{D_{mg}}$ *if* $0 \le \rho \le \rho^*$, *where* D_{mg} *is the WDB of DSCT with the* (σ_i, ρ_i)*-regulated general MUX, and we give its value in Remark 2.*
3. *When* K *is large enough, the ratio of the range* $[\rho^*, \frac{1}{K})$ *to the total range* $(0, \frac{1}{K})$ *is approximately given by* $\frac{\frac{1}{K}-\rho^*}{\frac{1}{K}} \approx \frac{5-\sqrt{21}}{2} \approx 0.21$.
4. *For any positive integer* n *such that* $\frac{1}{K} - \frac{1}{K^{n+1}} \ge \rho^*$, *we have* $\frac{D_{mg}}{\hat{D_{mg}}} \ge O(K^n)$ *whenever* $\bar{\rho} \in [\frac{1}{K} - \frac{1}{K^{n+1}}, \frac{1}{K})$.

Proof. 1) Suppose the longest multicast path (denoted as $< s^i \to r^i >$) in G^i is the one connecting the source s^i and the receiver r^i, where $s^i, r^i \in G^i$, and $s^i \ne r^i$. Assume that there are F forwarders on the path $< s^i \to r^i >$ that are denoted as the set of $\{\gamma_1^i, \ldots, \gamma_m^i, \ldots, \gamma_F^i\}(m \in [1, F])$ and $\gamma_m^i \in G^i$. The worst-case multicast delay in G^i with the $(\sigma_i^*, \rho_i, \lambda_i)$-regulated general MUX is the worst-case delay of any bit passing through $< s^i \to r^i >$ when s^i and all γ_m^i join in all the K groups. Then, the worst-case multicast delay bound $D_{mg}^{\hat{i}}$ in G^i is calculated by

$$D_{mg}^{\hat{i}} = \hat{D}_g^i(< s^i \to \gamma_1^i >) + \hat{D}_g^i(< \gamma_F^i \to r^i >) + \sum_{m=1}^{F-1} \hat{D}_g^i(< \gamma_m^i \to \gamma_{m+1}^i >),$$

where $\hat{D}_g^i(< s^i \to \gamma_1^i >)$, $\hat{D}_g^i(< \gamma_F^i \to r^i >)$, and $\sum_{m=1}^{F-1} \hat{D}_g^i(< \gamma_m^i \to \gamma_{m+1}^i >)$ refer to the WDBs between s^i and γ_1^i, γ_F^i and r^i, and γ_m^i and γ_{m+1}^i, respectively. According to Theorem 1, they are equal to

$$\sum_{i=1}^{K} \frac{\sigma_i^*}{1-\rho_i} + 2 \min_{1 \le i \le K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} + \max_{1 \le i \le K}\left\{\frac{(\sigma_i - \sigma_i^*)}{\rho_i}\right\}.$$

Hence, the worst-case delay $D_{mg}^{\hat{i}}$ of any bit passing through the DSCT tree in G^i is

$$D_{mg}^{\hat{i}} = (H_i - 1)\left[\sum_{i=1}^{K} \frac{\sigma_i^*}{1-\rho_i} + 2 \min_{1 \le i \le K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} + \max_{1 \le i \le K}\left\{\frac{(\sigma_i - \sigma_i^*)}{\rho_i}\right\}\right],$$

where H_i is the height bound of the DSCT tree in G^i. Actually, $H_i = m + 1$, where m is the number of forwarders in the longest path $< s^i \to r^i >$.

Considering the whole multigroup network, the worst case multicast delay occurs in the group with the highest DSCT tree. We have

$$\hat{D_{mg}} = \max_{1 \le i \le K}\{D_{mg}^{\hat{i}}\} = (\hat{H} - 1) \left[\sum_{i=1}^{K} \frac{\sigma_i^*}{1-\rho_i} + 2 \min_{1 \le i \le K}\left\{\frac{\sigma_i}{\rho_i(1-\rho_i)}\right\} + \max_{1 \le i \le K}\left\{\frac{(\sigma - \sigma_i^*)}{\rho_i}\right\}\right],$$

where $\hat{H} = \max_{1 \le i \le K}\{H_i\}$. □

The proofs of 1, 2, and 4 can be similarly established as the proof of Theorems 3 and 5 and, thus, are omitted here.

For the homogeneous flows in the multigroup network, we present Theorem 8.

Theorem 8. *Suppose there are* K *groups denoted as* $G^i (i \in [1, K])$ *in the regulated multigroup network, and each group has* n_i *end hosts that construct a DSCT tree. If one real-time flow exists in each group, and the flow is constrained by the rate function* R_i *such that* $R_i \sim (\sigma_0, \rho)$ *with the stability condition* $\rho \le \frac{1}{K} (\hat{K} \in [1, K])$ *at each end host that joins in* $\hat{K}$ *groups, then*

1. *The maximum worst case delay experienced by any bit passing through the DSCT tree with the* (σ, ρ, λ)*-regulated general mux is upper bounded by* $\hat{D_{mg}} = \frac{(\hat{H}-1)K\sigma}{1-\rho} + \frac{(\hat{H}-1)(\sigma_0 - \sigma)^+}{\rho} + \frac{2(\hat{H}-1)\lambda\sigma}{\rho}$, *where* $\hat{H} = \max_{1 \le i \le K}\{H_i\}$, *and* H_i *is the height bound of the DSCT tree in* G^i *that can be derived by Lemma 2.*
2. *If* $K \ge 2$ *is satisfied, then there exists a rate threshold* $0 < \rho^* < \frac{1}{K}$ *such that* $\hat{D_{mg}} \le D_{mg}$ *if* $\rho^* \le \bar{\rho} < \frac{1}{K}$, *and* $D_{mg} \le \hat{D_{mg}}$ *if* $0 < \bar{\rho} \le \rho^*$, *where* D_{mg} *is the WDB of any bit passing through the DSCT tree with the* (σ, ρ) *regulator, and we give its value in Remark 2.*
3. *When* K *is large enough, the ratio of the range* $[\rho^*, \frac{1}{K})$ *to the total range* $(0, \frac{1}{K})$ *is approximately given by* $\frac{\frac{1}{K}-\rho^*}{\frac{1}{K}} \approx 2 - \sqrt{3} \approx 0.27$.
4. *For any positive integer* n *such that* $\frac{1}{K} - \frac{1}{K^{n+1}} \ge \rho^*$, *we have* $\frac{D_{mg}}{\hat{D_{mg}}} \ge O(K^n)$ *whenever* $\bar{\rho} \in [\frac{1}{K} - \frac{1}{K^{n+1}}, \frac{1}{K})$.

Remark 2. Based on [15, (13)], suppose any bit passing through the network with K groups are regulated by the general muxes. The general mux at each end host has $\hat{K}$ input links (that is, the end host joins in $\hat{K}$ groups), and the rate functions for the flows in the $\hat{K}$ groups are given by R_i such that $R_i \sim (\sigma_i, \rho_i)(R_i \sim (\sigma_0, \rho))$, $i \in [1, \hat{K}]$, with the stability condition $\sum_{i=1}^{\hat{K}} \rho_i \le 1 (\rho \le \frac{1}{K})$ at each end host that joins in $\hat{K}(\hat{K} \in [1, K])$ groups. Then, the maximum delay of the data bit is upper bounded by $D_{mg} = \frac{(\hat{H}-1)\sum_{i=1}^{K}\sigma_i}{1-\sum_{i=1}^{K}\rho_i} (D_{mg} = \frac{(\hat{H}-1)K\sigma_0}{1-K\rho})$, where $\hat{H}$ is the maximum value of the DSCT tree height bounds of K groups.

6 Simulation Evaluation

In this section, we use the simulations to evaluate the WDBs of network communications with and without our *adaptive control algorithm*, respectively. We have done two groups of simulations in ns-2 [22] and run them on a group of SUN SOLARIS workstations.

6.1 Simulation 1

In the first group of simulations, we observe the WDB performances of the single $(\sigma, \rho, \lambda)/(\sigma, \rho)$-regulated end

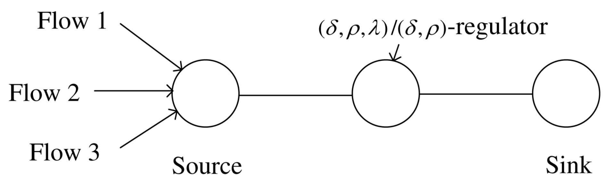


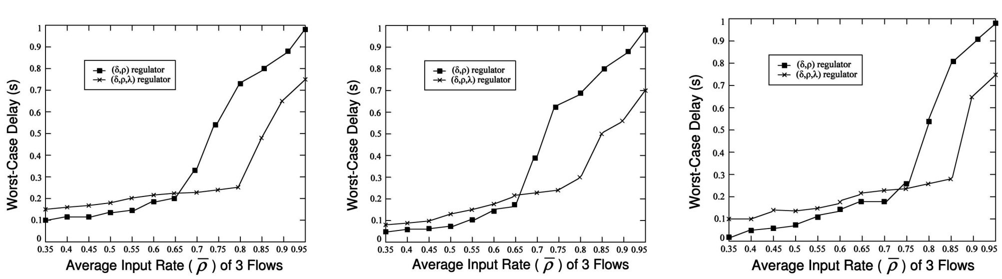
Fig. 3. The simulation topology with only one $(\sigma, \rho, \lambda)/(\sigma, \rho)\text{-regulated}$ end host.

host. Fig. 3 shows the simulation topology. The source is fed with three real-time flows that are going to transmit to the sink. The intermediate node is equipped with the $(\sigma, \rho, \lambda)/(\sigma, \rho)\text{-regulated}$ general muxes, respectively. Two types of real-time streams are employed: audio streams with 64 kilobits per second and MPEG-1 video streams with 1.5 megabits per second. We compare the WDB performances of the (σ, ρ, λ) regulator and the (σ, ρ) regulator with three video streams, three audio streams, and three heterogeneous streams (one video and two audio streams), respectively.

Fig. 4a illustrates the worst case delay performances when there are three 64-Kbps audio streams that pass through the network shown in Fig. 3. The simulation results meet our theoretical analysis. The cross point of the two curves is 0.66; that is, the input rate threshold in this simulation is 0.66. When $\bar{\rho} < 0.66$, the worst-case delays of the packets passing through the (σ, ρ, λ) regulator are longer than the worst-case delays of the packets passing through the (σ, ρ) regulator. Otherwise, the worst-case delays with the (σ, ρ, λ) regulator are shorter than the ones with the (σ, ρ) regulator. The rate threshold difference (between the simulation result and the theoretical analysis) is because our theoretical analysis does not take into account the fluctuation of the network throughput in the simulation. The throughput fluctuation is mainly caused by the following reasons: 1) the cluster size is a random integer between k and $3k - 1$, which possibly makes the same end host have different child members in different multicast schemes and 2) the audio and video streams in the simulation are all variable bit rate (VBR) flows. The transmission rates of the VBR flows are changing over time, which causes the fluctuation of the throughput. When the number of VBR flows increases, the fluctuation of the network throughput becomes large. Also, it can be seen in Fig. 4a that when $\bar{\rho} \geq 0.66$, the maximum worst-case delay improvement of the (σ, ρ, λ) regulator over the (σ, ρ) regulator is at $\bar{\rho} = 0.8$ and has the value of $\frac{0.72}{0.26} \approx 2.8$. According to Theorem 6, we can derive $n \approx 1$ from the simulation parameter $K = 3$. Fig. 4b illustrates the WDB performances of three homogeneous video streams. The rate threshold of three flows is 0.67, which is a little less than the theoretical result 0.73 for the fluctuation of the network throughput. The maximum improvement in the worst-case delays of the (σ, ρ, λ) regulator over the (σ, ρ) regulator is at $\bar{\rho} = 0.8$ and has the value of $\frac{0.72}{0.26} \approx 2.82$. According to Theorem 6, we can also derive $n \approx 1$ from the simulation parameter $K = 3$. Fig. 4c gives the comparison of the worst case delay performance of heterogeneous real-time streams in the network. It can be seen that the input rate threshold of three flows is 0.74, which is a little less than the theoretical value of 0.79 in Theorem 3. When $\bar{\rho} \geq 0.74$, the worst-case delays with the (σ, ρ, λ) regulator are much shorter than the ones with the (σ, ρ) regulator. The maximum improvement in the worst-case delay is at $\bar{\rho} = 0.85$ and with the value of $\frac{0.85}{0.27} \approx 3.15$, which meets the theoretical results in Theorem 5 when $n = 1$.

6.2 Simulation 2

In the second group of simulations, we observe the worst-case delay performances of real-time streams in the multigroup network. There are 665 end hosts in the network who join in three groups. Fig. 5 shows the backbone network topology. The 665 group members directly or indirectly, through some intermediate network components (for example, the hubs), attach to the routers in the backbone network and are with the $(\sigma, \rho, \lambda)/(\sigma, \rho)\text{-regulated}$ general muxes. Each group has one real-time flow. That is, each end host needs to serve three real-time flows. Also, there are two types of simulation streams: 64-Mbps audio streams and 1.5-Mbps MPEG-1 video streams in the multigroup network. In this group of simulations, we compare the WDB performances under three EMcast schemes for Nice is the Internet Cooperative Environment (NICE) and DSCT multicast trees, respectively: the *capacity-aware multicast tree*, the *multicast tree with the* (σ, ρ) *regulator*, and the *multicast tree with the* (σ, ρ, λ) *regulator*. The traffic pattern is three video streams, three audio streams, and three heterogeneous streams (one video and two audio streams), respectively.



(a) (b) (c)

Fig. 4. The worst-case delay performances when there are (a) three 64-Kbps audio streams, (b) three 1.5-Mbps video streams, and (c) one 1.5-Mbps video stream and two 64-Kbps audio streams in the network.

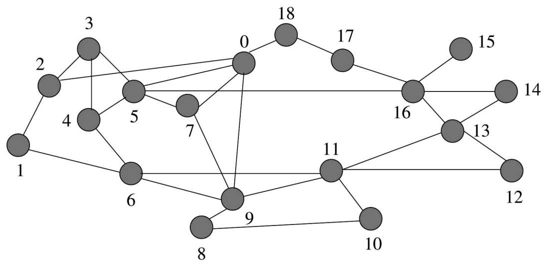


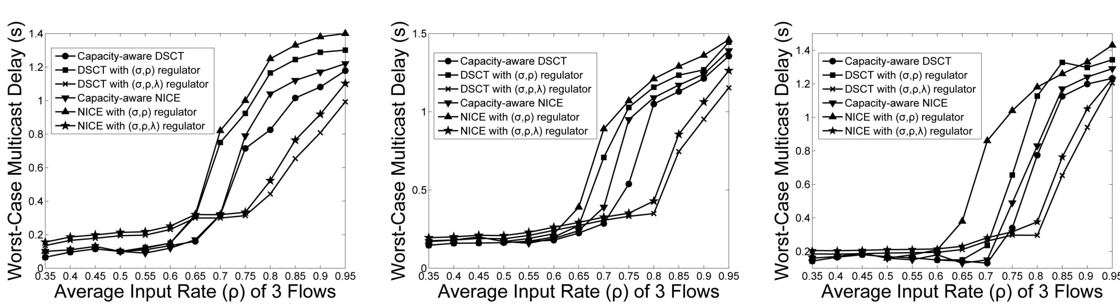
Fig. 5. The backbone network topology in the simulations.

Fig. 6a illustrates the worst case delay performances of the *capacity-aware multicast tree*, the *multicast tree with the* (σ, ρ) *regulator*, and the *multicast tree with the* (σ, ρ, λ) *regulator* for NICE and DSCT, respectively, when each of the three groups is fed with the same 64-Kbps audio stream. In the figure, we can see that the *capacity-aware DSCT* can achieve shorter delay performances than the *DSCT with the* (σ, ρ) *regulator*. Moreover, when $\bar{\rho} \geq 0.7$, the *DSCT with the* (σ, ρ, λ) *regulator* achieves the best delay performances in the three multicast schemes. Compared to the *DSCT with the* (σ, ρ) *regulator*, the rate threshold of the three flows in the simulation is 0.65, which is a little less than the theoretical value of 0.73 in Theorem 8. Furthermore, the maximum improvement in the worst-case delays of the *DSCT with the* (σ, ρ, λ) *regulator* over the *DSCT with the* (σ, ρ) *regulator* is at $\bar{\rho} = 0.75$ and has the value of $\frac{0.95}{0.27} \approx 3.52$, which meets the theoretical results in Theorem 8 when $n = 1$. Also, in this figure, we can see that the curves of the NICE tree with the (σ, ρ) regulator, the NICE tree with the (σ, ρ, λ) regulator, and the capacity-aware NICE tree have the similar comparison trend as the ones of the DSCT tree schemes. The NICE tree with the (σ, ρ, λ) regulator achieves shorter worst-case delay performances than the capacity-aware NICE tree when the average transmission rate becomes high. Furthermore, the curves show that the DSCT tree achieves shorter worst-case delays than the NICE tree when they employ the same traffic control schemes. The results meet our analysis in [14]. It is mainly because DSCT employs the hosts' location knowledge to build up the multicast architecture.

Fig. 6b shows the worst-case multicast delay performances of the video streams. The *capacity-aware DSCT* achieves shorter delay performances than the *DSCT with the* (σ, ρ) *regulator*, and when $\bar{\rho} \geq 0.7$, the *DSCT with the* (σ, ρ, λ) *regulator* achieves the shortest delay performances in the three multicast schemes. As for the comparison of the *DSCT with the* (σ, ρ, λ) *regulator* to the *DSCT with the* (σ, ρ) *regulator*, the simulation rate threshold of the three flows is 0.65, and the maximum worst-case multicast delay improvement of the *DSCT with the* (σ, ρ, λ) *regulator* over the *DSCT with the* (σ, ρ) *regulator* is at $\bar{\rho} = 0.8$ and with the value of $\frac{1.18}{0.32} = 3.69$. Similar to the curves in Fig. 6a, in homogeneous video communications, the NICE tree with the (σ, ρ, λ) regulator achieves 26 shorter worst-case delay performances than the capacity-aware NICE tree when the average transmission rate becomes high. Also, the worst-case delays of the NICE tree in each traffic control scheme are longer than the corresponding ones of the DSCT tree.

Fig. 6c gives the worst-case delay performance comparison when one group is fed with the 1.5-Mbps video stream, and each of the other groups is fed with the 64-Kbps audio stream. The simulation results also tell us that the rate threshold of the three flows is 0.735, which is a little less than the theoretical result of 0.79 in Theorem 7 because of the network throughput fluctuation in the practical network. Moreover, the maximum worst-case delay improvement of the *DSCT with the* (σ, ρ, λ) *regulator* over the *DSCT with the* (σ, ρ) *regulator* is at $\bar{\rho} = 0.8$ and has the value of $\frac{1.15}{0.27} \approx 4.26$. Also, the comparison of the NICE tree schemes and the DSCT tree schemes shows the similar trend as the one in Fig. 6a.

Table 1 gives the comparison of the multicast tree layer numbers when the three groups are with the homogeneous audio streams. Data in the table show that the DSCT with the (σ, ρ, λ) regulator achieves shorter delay performances without increasing the tree height. However, the height of the capacity-aware DSCT increases with the increment of the average input rate. Tables 2 and 3 are the comparison of the multicast tree layer numbers when the three groups are with the homogeneous video streams and heterogeneous streams, respectively. Similar to the results in Table 1, data in these two tables prove that the (σ, ρ, λ) regulator reduces the worst-case delay without increasing the tree height.



(a) (b) (c)

Fig. 6. The worst-case delay performances when each of the three groups is fed with the same (a) 64-Kbps audio streams, (b) 1.5-Mbps video streams, and (c) one 1.5-Mbps video stream and two 64-Kbps audio streams.

TABLE 1
Comparison of Tree Layer Numbers (Three Groups with Homogeneous Audio Streams)

$\bar{\rho}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
Capacity-aware DSCT	5	5	5	6	6	6	7	8	7	8	8	9	9
DSCT with (σ, ρ, λ) regulator	6	6	6	6	6	6	6	6	6	6	6	6	6

TABLE 2
Comparison of Tree Layer Numbers (Three Groups with Homogeneous Video Streams)

$\bar{\rho}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
Capacity-aware DSCT	6	7	7	8	8	8	9	9	9	10	9	10	10
DSCT with (σ, ρ, λ) regulator	7	7	7	7	7	7	7	7	7	7	7	7	7

TABLE 3
Comparison of Tree Layer Numbers (Three Groups with Heterogeneous Streams)

$\bar{\rho}$	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
Capacity-aware DSCT	5	5	6	6	6	7	7	7	8	8	8	9	9
DSCT with (σ, ρ, λ) regulator	6	6	6	6	6	6	6	6	6	6	6	6	6

7 Conclusion

In this paper, we addressed the problem of decreasing the WDB for EMcast when the group members are in the face of having no enough capacities to output the simultaneous input traffic. We presented a novel *adaptive control algorithm*. Based on the instantaneous network situations, the algorithm adaptively employs the (σ, ρ) regulator under the normal traffic load situation and the (σ, ρ, λ) regulator under the heavy traffic load situation to control the traffic output at each end host. The (σ, ρ, λ) regulator adopts two states (*on* and *off*) to assign the output of the simultaneous heavy input flows in turn without increasing the multicast tree height. Through using *network calculus*, we proved a set of theorems on the *input rate threshold* ρ^*, above which the (σ, ρ, λ) regulator benefits the shorter delay performances, the *WDB* for the (σ, ρ, λ) regulator achieves, and the *improvement of the WDB* of the (σ, ρ, λ) regulator over the (σ, ρ) regulator for single end host and EMcast with homogeneous and heterogeneous flows, respectively.

We then study our algorithm in the simulation environments. We ran two groups of simulations with the single regulated end host topology and the EMcast topology, respectively. We observed the worst-case delay improvement of the (σ, ρ, λ) regulator over the (σ, ρ) regulator and the rate threshold. The simulation results meet our theoretical analysis. Therefore, the possible bottleneck in multigroup network can be avoided without increasing the lengths of the multicast paths. When the flow coexists with other traffic, the number of input traffic at the end host is changed, and the flows' average input rate may be increased or decreased for the changed traffic load. Such change influences the values of each flow's working period and vacation period. However, we think that the same process of the adaptive control algorithm can be implemented to control the traffic and its coexisted flows when the traffic priority is ignored. When the traffic priority is considered, we should extend our algorithm to deal with the flows with different priorities. For example, we can add new parameters into the (σ, ρ, λ) regulator to enable it to recognize and process flows with different priorities. In the next step, we also propose to test our algorithm in the real-world network environment (for example, PlanetLab). Furthermore, to study the algorithms on other QoS requirements (for example, error control and packet loss) in multicast communications through theorems and simulations is our nearly future work.

Acknowledgments

The authors are grateful to the editor and reviewers for their invaluable and helpful comments. The authors thank Professor Hanxing Wang for help with mathematics and Mr. Mark O'Brien for help with English. This work was supported by the Embark Postdoctoral Fellowship of Ireland with the funding code 501-et-504 4890.

REFERENCES

[1] Y.H. Chu, S. Rao, S. Seshan, and H. Zhang, "A Case for End System Multicast," *Proc. ACM Int'l Conf. Measurement and Modeling of Computer Systems (SIGMETRICS '00)*, pp. 1-12, June 2000.

[2] P. Francis, "Yoid: Extending the Internet Multicast Architecture," http://www.aciri.org/yoid/docs/index.html, Apr. 2000.

[3] Y. Chawathe, "Scattercast: An Architecture for Internet Broadcast Distribution as an Infrastructure Service," PhD dissertation, Univ. of California, Berkeley, Aug. 2000.

[4] D. Pendarakis, S. Shi, D. Verma, and M. Waldvogel, "ALMI: An Application Level Multicast Infrastructure," *Proc. Third Usenix Symp. Internet Technologies and Systems (USITS '01)*, pp. 49-60, Mar. 2001.

[5] S. Shi and J. Turner, "Routing in Overlay Multicast Networks," *Proc. IEEE INFOCOM '02*, pp. 1200-1208, June 2002.

[6] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker, "Application-Level Multicast Using Content-Addressable Networks," *Proc. Third Int'l Workshop Networked Group Comm. (NGC '01)*, pp. 14-29, Nov. 2001.

[7] B. Zhang, S. Jamin, and L. Zhang, "Host Multicast: A Framework for Delivering Multicast to End Users," *Proc. IEEE INFOCOM '02*, pp. 1366-1375, June 2002.

[8] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, "Scalable Application Layer Multicast," *Proc. ACM SIGCOMM '02*, pp. 205-217, Aug. 2002.

[9] M. Castro, P. Druschel, A.M. Kermarrec, and A. Rowstron, "SCRIBE: A Large-Scale and Decentralized Application-Level Multicast Infrastructure," *IEEE J. Selected Areas in Comm.*, vol. 20, no. 8, pp. 1489-1499, Oct. 2002.

[10] J. Jannotti, D.K. Gifford, K.L. Johnson, M.F. Kaashoek, and J.W. O'Toole, Jr., "Overcast: Reliable Multicasting with an Overlay Network," *Proc. Fourth Usenix Symp. Operating Systems Design and Implementation (OSDI '00)*, pp. 194-210, Oct. 2000.

[11] H. Deshpande, M. Bawa, and H. Garcia-Molina, "Streaming Live Media over Peers," *Proc. First Workshop Hot Topics in Networks (HotNets-I)*, pp. 1-14, Oct. 2002.

[12] S. Banerjee, C. Kommareddy, B.B.K. Kar, and S. Khuller, "Construction of an Efficient Overlay Multicast Infrastructure for Real-Time Applications," *Proc. IEEE INFOCOM '03*, pp. 1521-1531, 2003.

[13] A. Riabov and L.Z. Zhen Liu, "Overlay Multicast Trees of Minimal Delay," *Proc. 24th Int'l Conf. Distributed Computing Systems (ICDCS '04)*, pp. 654-661, Mar. 2004.

[14] W. Tu and W. Jia, "A Scalable and Efficient End Host Multicast for Peer-to-Peer Systems—DSCT," *Proc. 47th Ann. IEEE Global Telecomm. Conf. (Globecom '04)*, pp. 967-971, 2004.

[15] R. Cruz, "A Calculus for Network Delay, Part I: Network Elements in Isolation," *IEEE Trans. Information Theory*, vol. 37, no. 1, pp. 114-131, Jan. 1991.

[16] R. Cruz, "A Calculus for Network Delay, Part II: Network Analysis," *IEEE Trans. Information Theory*, vol. 37, no. 1, pp. 131-141, Jan. 1991.

[17] D. DeLucia and K. Obraczka, "Multicast Feedback Suppression Using Representatives," *Proc. IEEE INFOCOM '97*, pp. 463-470, Apr. 1997.

[18] H.A. Wang and M. Schwartz, "Achieving Bounded Fairness for Multicast Traffic and TCP Traffic in the Internet," *Proc. ACM SIGCOMM '98*, pp. 81-92, 1998.

[19] J. Widmer and M. Handley, "Extending Equation-Based Congestion Control to Multicast Applications," *Proc. ACM SIGCOMM '01*, Aug. 2001.

[20] I. Rhee, N. Ballaguru, and G.N. Rouskas, "MTCP: Scalable TCP-Like Congestion Control for Reliable Multicast," *Proc. IEEE INFOCOM '99*, pp. 1265-1273, Mar. 1999.

[21] J. Golestani and K. Sabnani, "Fundamental Observations on Multicast Congestion Control in the Internet," *Proc. IEEE INFOCOM '99*, pp. 990-1000, Mar. 1999.

[22] *NS Notes and Documentation.* UC Berkeley, LBL, USC/ISI, and Xerox PARC, Oct. 20,1999.

[23] J. Turner, "New Directions in Communications," *IEEE Comm. Magazine*, vol. 24, no. 10, pp. 8-15, Oct. 1986.

[24] M. Sidi, W.Z. Liu, I. Cidon, and I. Gopal, "Congestion Control through Input Rate Regulation," *Proc. IEEE Global Telecomm. Conf. (Globecom '89)*, pp. 1764-1768, Nov. 1989.

[25] A.E. Echberg, D.T. Luan, and D.M. Lucantoni, "An Approach to Controlling Congestion in ATM Networks," *Int'l J. Digital Analog Comm. System*, vol. 3, pp. 199-209, 1990.

[26] W. Jia, H. Wang, M. Tang, and W. Zhao, "Effective Delay Control for High Rate Heterogeneous Real-Time Flows," *Proc. 23rd IEEE Int'l Conf. Distributed Computing Systems (ICDCS '03)*, pp. 367-375, May 2003.

[27] M. Rodriguez-Perez, M. Fernandez-Veiga, C. Lopez-Garcia, J.C. Lopez-Ardao, and S. Herreria-Alonso, "An Open-Loop Multicast Layered Congestion Protocol for Real-Time Multimedia Transmission," *Proc. IEEE Global Telecomm. Conf. (Globecom '03)*, pp. 2855-2859, Dec. 2003.

[28] L. Rizzo, "Pgmcc: A TCP-Friendly Single-Rate Multicast Congestion Control Scheme," *Proc. ACM SIGCOMM '00*, 2000.

[29] J. Widmer and M. Handley, "Extending Equation-Based Congestion Control to Multicast Applications," *Proc. ACM SIGCOMM '01*, Aug. 2001.

[30] A. Chaintreau, F. Baccelli, and C. Diot, "Impact of Network Delay Variation on Multicast Sessions Performance with TCP-Like Congestion Control," *IEEE Trans. Networking*, 2002.

[31] Y. Amir and C. Danilov, "Reliable Communication in Overlay Networks," *Proc. 37th IEEE Int'l Conf. Dependable Systems and Networks (DSN '03)*, 2003.

[32] G.U. Keller and E.W. Biersack, "A Congestion Control Model for Multicast Overlay Networks and Its Performance," *Proc. Fourth Int'l Workshop Networked Group Comm. (NGC '02)*, Oct. 2002.

[33] Y. Amir, B. Awerbuch, C. Danilov, and J. Stanton, "Global Flow Control for Wide Area Overlay Networks: A Cost-Benefit Approach," *IEEE OpenArch*, 2002.

[34] G.I. Kwon and J.W. Byers, "ROMA: Reliable Overlay Multicast with Loosely Coupled TCP Connections," *Proc. IEEE INFOCOM '04*, 2004.

[35] J. Lu, Q. Ruan, and R. Ni, "A Scalable Overlay Multicast Congestion Control for Multimedia Streaming," *Proc. 30th IEEE Conf. Local Computer Networks (LCN '05)*, Nov. 2005.

[36] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "TCP Friendly Rate Control (TFRC): Protocol Specification," RFC 3448, Network Working Group, Jan. 2003.

Wanqing Tu received the PhD degree from the Department of Computer Science, City University of Hong Kong, in 2006. She is a postdoctoral researcher of computer science at University College Cork (UCC), Ireland. Her research interests include quality of service (QoS), overlay networks, wireless mesh networks, end host multicast, and distributed computing. She is a member of the IEEE. She received the Embark Postdoctoral Fellowship of the Irish government and one Best Paper Award.

Cormac J. Sreenan (M'93) received the PhD degree in computer science from Cambridge University. He is a professor of computer science at University College Cork (UCC), Ireland. Prior to joining UCC in 1999, he was a research staff member at AT&T Laboratories—Research, Florham Park, New Jersey, and at Bell Laboratories, Murray Hill, New Jersey. At UCC, he directs the Mobile and Internet Systems Laboratory (MISL), which is a group of more than 10 research staff and students with research activity in multimedia and wireless networking and systems, including wireless sensor networks and media content distribution. He is currently on the editorial boards of the *ACM/Springer Multimedia Systems Journal* and the *Elsevier Computer Networks Journal*. In the past, he has served as a guest editor for the *Communications of the ACM*, the *IEEE Journal on Selected Areas in Communications*, and the *IEEE Wireless Communications Magazine*. In 2007, he will host and cochair the Proceedings of the Workshop on Embedded Networked Sensors (EmNets) and the Proceedings of the Conference on Mobile and Wireless Communication Networks (MWCN). He is a fellow of the British Computer Society (BCS) and a member of the IEEE.

Weijia Jia received the BSc, MSc, and PhD degrees, all in computer science, in 1982, 1984, and 1993, respectively, from Center South University, China, and Polytechnic Faculty of Mons, Belgium. He is an associate professor in the Department of Computer Science, City University of Hong Kong. His research interests include computer network, distributed systems, and multicast and anycast QoS routing protocols for Internet and wireless communications. He is a member of the IEEE.