Service Curve Allocation For End-to-end Delay Requirement

Yen-Ping Chu
Department of Applied Mathematics
National Chung-Hsing University
Taichung, Taiwan
E-mail:ypchu@amath.nchu.edu.tw

Chin-Hsing Chen
Department of Management Information System
Chungtai Institute of Health Sciences and Technology
Taichung, Taiwan
E-mail:chchen@chtai.ctc.edu.tw

Abstract

An application with guaranteed service only cares about whether or not the network can satisfy its performance requirement, such as end-to-end delay. However, the network wants to achieve the high utilization and performance guarantee simultaneously. If the end-to-end delay provided by the network can be allocated properly to each switching node, then the network resources can get a better utilization. Conventionally, the delay is allocated equally to each switching node along the path that the connections pass through, referred to as Even division policy. The advantage of this policy is easy to implement. However, we can not understand how this policy will affect the network utilization. In this paper, we proposed an allocation scheme called MaxMin allocation to improve network utilization. Under SCED(Service Curve Earlist Deadline) scheduling policy, we reduce the service curve allocation to end-to-end delay allocation, and using MaxMin delay allocation policy, to promote the network utilization. From simulations, we find that the performance of MaxMin policy is better than Even division policy.

Keyword: guaranteed service, scheduling, local QOS allocation, service curve.

1 Introduction

Current high-speed networks must provide a variety of network service classes to accommodate the diverse QoS requirements of multimedia applications. Guaranteed-service is applied to application with diverse traffic characteristic and different QoS requirements. Applications with guaranteed service would like to acquire an end-to-end delay bound from the network. Usually, an end user is only concerned with the QoS requirements on an end-to-end basis and does not care about the QoS of the local switching node. One important parameter of QoS is the delay through the network experienced by the applications. But, the end-to-end delay bounds associated with current papers always be overly conservative, limiting the utilization of the network resource. A very important objective of integrated service networks is to be able to support the maximum allowable connections, while guaranteeing the service for each connection. The scheduling and admission control mechanisms are both the important policies in achieving this objective. When a new user wants to enter the network, it has to specify the types of service that the network must provide for the user. The network will do an admission control test to decide whether the user can be supported without disrupting the service for existing users. Once the user is accepted, the network will make use of a scheduling policy at each switch in the network to allocate resources so that each user can receive its requested service. The scheduling policy should support applications with diverse QoS requirements and should provide a user with a flexible means for specifying the service that best matches the needs of the user. Furthermore, the policy should admit as many connections as possible without violating the delay quarantees for each connection, and a connection should be protected by the scheduling policy from traffic fluctuation in the network and misbehaving users.

Service curve proposed by Cruz[5] is a means for characterizing the service provided for a connection by a network. This service can be translated into delay guarantees for the connection provided the traffic characteristic of the connection at the entrance of the network is known. A service curve lower bounds the output traffic of the switch in some interval, and if the service is bounded in this way, it had been shown that the connection receives some delay guarantees. Our proposed method is based on service curves. For promoting network utilization, we propose a scheme for allocating service curve at the servers in a tandem network such that a prespecified network service curve is guaranteed and the excess bandwidth is maximized.

This paper is organized as follows. In section 2, we describe some basic ideas for our scheme. The concepts of service curves and SCED scheduling are reviewed. Section 3 shows our proposed scheme. The simulation results are presented in section 4. Section 5 concludes our discussions.

2 Basic ideas

Our framework is based on Rate Controlled Service Discipline (RCSD)[3]. A connection specifies a burstiness constraint that bounds the amount of generated traffic, and the network elements employ shapers and schedulers to protect individual connection from one another. We use a scheduling policy called SCEDs[4][6], which service a connection according to the service curve specification. This scheme has the following features: flexibility in allocating bandwidth and delay for different connections, efficiency in admitting as many connections as possible, and simplicity in implementation. The following sections will introduce some definitions and results about service curves and SCED scheduling.

2.1 Serive curves

Consider a virtual circuit connection that passes through M packet switches in tandem. Let R^{m-1} describe the traffic entering switch m, and the traffic departing switch m feeds

switch $m+1$, for $1 \le m \le M$.

Define $R_{in} = R^0$, $R_{out} = R^M$,

$d^m[t] = \min\{\Delta : \Delta \ge 0 \text{ and } R^0[1,t] \le R^m[1,t+\Delta]\}$, and

$B^m[t] = R^{m-1}[1,t] - R^m[1,t]$ where $d^m[t]$ is the virtual delay of the connection through the first m switches, $B^m[t]$ is the backlog at the end of time t at switch m. The following definitions and theorems were proposed by Cruz etc.[5]

Definition 1: (Arrival constraints) Given a nondecreasing function $b(\cdot)$, we say that R^m is b-smooth if $R^m[s+1,t] \le b(t-s)$ for all s and t satisfying $s \le t$. Specially, if $b(x) = \sigma + \rho x$, then we say that R^m is (σ, ρ)-smooth.

Definition 2: (End-to-end service constraints) The network is said to guarantee a service curve of $S_{net}(\cdot)$ if for all t, there exits a slot $s \le t$ such that $B^1[s] = 0$, and $R_{out}[1,t] - R_{in}[1,s] \ge S_{net}(t-s)$.

Based on the above definitions, Cruz etc. have proven the following theorems.

Theorem 1: (End-to-end service) Suppose that each network element m guarantees the given connection a service curve of $S^m(\cdot)$. Then the network gurantees a service curve of $S_{net}(\cdot)$, where

$$S_{net}(x) = \min\{\sum_{m=1}^{M} S^m(x^m) : x^m \ge 0, \text{and } \sum_{m=1}^{M} x^m = x\}.$$

This theorem means that the network service curve can be computed by computing the service curve of each independent switch, moreover, the computing of network service curve doesn't depend on the ordering of the switch.

Theorem 2: (Upper bound on end-to-end delay) Suppose that the network guarantees a service curve of $S_{net}(\cdot)$ and suppose that the input traffic to the system, R_{in} is b-smooth. Then the end-to-end delay $d^M[t]$ is upper bounded by

$d^M[t] \le \max_{k:k\ge 1} \min\{\Delta : \Delta \ge 0 \text{ and } b(k) \le S_{net}(k+\Delta)\}$.

From these results, we find that bounds on end-to-end delay can be obtained in terms of service curves and burstiness constraints on arriving traffic.

2.2 SCEDs scheduling

In this section, we will introduce the SCED(Service Curve Earliest Deadline) scheduling proposed by H. Sariowan.
Consider a switch there are N connections passing through it. For nth connection, let R_n^{m-1} and R_n^m describe the input and output traffic for the switch m. $B_n^m[t]$ is the backlog for connection n in the switch at the end of time t. And let $S_n^m(\cdot)$ be a given nondecreasing nonnegative function for $1 \le n \le N$. The following scheduling policy can guarantee each connection n a service curve of $S_n^m(\cdot)$.

Each packet arriving from connection n at time t is assigned a deadline $D_n^m[t]$ according to

$D_n^m[t] = \min\{\Delta : \Delta \ge t \text{ and } Z_n^m(\Delta; t-1) \ge R_n^m[1,t]\}$,

where

$Z_n^m(k;t) = \min\{R_n^m[1,u] + S_n^m(k-u) : 0 \le u \le t \text{ and } B_n^m[u] = 0\}$.

The policy serves packets in the increasing order of their deadlines. H. Sariowan have shown that [5] if packets are scheduled using this policy and each packet departs no later than its assigned deadline, then the connection is guaranteed a service curve.

Suppose a connection generates traffic being (σ, ρ)-smooth, and requires that the end-to-end delay is at most $d_{\max}$. Assume the switch uses SCED scheduling, then by Theorem 2 with $b(x) = \sigma + \rho x$, this delay guarantee can be made if the connection is allocated the network service curve S_{net}, where

$$S_{net} = \begin{cases} 0 & \text{if } 0 < t \le d_{\max} - 1 \\ \sigma + \rho(t - d_{\max}) & \text{if } t \ge d_{\max} \end{cases} \quad (1)$$

This means that we can find a network service curve for a connection according to its delay requirement and traffic constraint.

3 Service curve allocation

In this section, we will propose a scheme for allocating service curve at the servers in a tandem network such that the network service curve is quaranteed and the network utilization can be promoted. In the following, we first introduce our network model, and then show the allocation of service curve.

3.1 Network model

We consider a model consisting of M network elements in tandem. Each network element could be a packet switch along the route of a given connection. And we assume that these switching nodes are taken to a set of source-destination pairs. The input traffic envelope for a connection n is (σ_n, ρ_n)-smooth. And the delay requirement for this connection is d_n. For node m, the scheduling policy is SCED, and the output link capacity is R^m. The switch architecture is described in Figure 1.

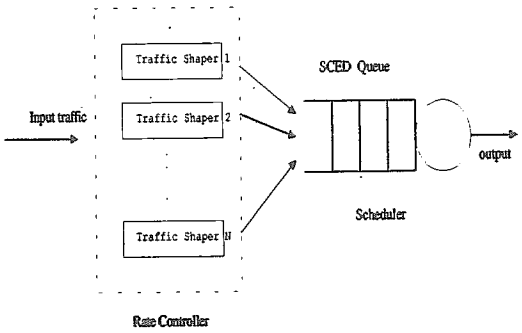

Figure 1 The Architecture of Switching Node

3.2 Allocation of service curve

Define P to be the vector $\{(\alpha_k, \beta_k)\}_{k=1}^{K}$. Assume that $\beta_1 > \beta_2 > \cdots > \beta_k > 0$. And assume the following inequality is satisfied:

$$1 \le \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2} < \cdots < \frac{\alpha_k - \alpha_{k-1}}{\beta_{k-1} - \beta_k}.$$

A CPL(Concave Piece-wise Linear) curve with parameter P, denoted by S^P is defined to be

$$S^P = Max\{0, \min_{k=1\cdots K}\{\alpha_k + \beta_k t\}\}.$$

$S^P = Max\{0, \sigma + \rho t\}$ is a special case with k=1, and parameter vector being P$\{(\sigma, \rho)\}$. Cruz etc.[6] proved the following theorem.

Theorem 3: (Allocation of service curve) Suppose a session desires CPL network service curve guarantee $S_{net}(\cdot)$ with parameter vector P$\{(\alpha_k, \beta_k)\}$. If at server m, m=1,...,M, the session is guaranteed CPL service curves $S^{-}(t;d)$ where

$$S^{-} = \begin{cases} 0 & \text{if } 0 < t \le d_{\max} - 1 \\ \left\lfloor S_{net}^{P}(t + d - d^m) \right\rfloor & \text{if } t \ge d_{\max} \end{cases}$$

and

$$d = \min\{\Delta : \Delta \ge 0 \text{ and } S_{net}(\Delta) \ge 1\},$$

and $d^m, m = 1, \cdots, M$ are arbitrarily assigned nonnegative real numbers which satisfy $\sum_{k=1}^{M} d_k \le d$, then the session is guaranteed network service curve $S_{net}(\cdot)$.

d is the end-to-end delay in Theorem 3.

The allocation of service curves at different servers is not unique. The network can assign any parameter d^m at server m, as long as the admission control condition is satisfied. Some optimum design criteria must be formulated. In the network view, the more the available bandwidth, the more allowable connections can be admitted to enter the network. Hence, we will propose a scheme to assign d^m at server m such that the excess bandwidth in the network can be maximized.

Because our input traffic for connection n is (σ_n, ρ_n)-smooth, and the delay request is d_n, from (1), the network service curve for this connection S_{net} can be

$$S_{net}(t) = \begin{cases} 0 & \text{if } 0 < t \le d_n - 1 \\ \sigma_n + \rho_n(t - d_n) & \text{if } t \ge d_n \end{cases}$$

In the following, we will present a scheme for allocating service curves at servers in a tandem network such that the connection is guaranteed the network service curve $S_{net}(t)$. The objective of our allocation is to maximize the excess bandwidth for the network. Using Theorem 3, we can allocate the service curve at server m as

$$\begin{aligned} S^m(t; d_n) &= S_{net}(t + d_n - d_m) \\ &= \sigma_n + \rho_n(t + d_n - d^m - d_n) \\ &= \sigma_n + \rho_n(t - d^m) \end{aligned}$$

where $t \ge d^m$ and $\sum_{m=1}^{M} d^m \le d_n$.

This implies that the allocation of service curves at the different servers can be reduced to assign parameter d^m at each server m. Let we specify

$$d^m = \frac{L}{g_n^m} + \frac{L}{R^m}, \qquad (2)$$

where g_n^m is the service rate of connection n at node m and L is the maximum packet size in the network. Therefore, g_n^m must be greater than or equal to ρ_n, or it will result to infinite delay, and be less than or equal to the residual bandwidth for the link of node m. Because our purpose is to assign d^m at server m such that the sum of the residual bandwidth of the network can be maximized, we can formulated our problem as follows:

Assume that the network contains M nodes labeled as 1,2,...,M respectively, and consists of a single source-destination pair of nodes between which connections are setup. And assume that there have been n-1 connections in the route. When a new connection n with (σ_n, ρ_n)-smooth wants to enter the network, and its delay requirement is d_n. We want to allocate the service curve at server m such that

$$\sum_{m=1}^{M}[(R^m - \sum_{i=1}^{n-1} g_i^m) - g_n^m] \text{ is maximized,}$$

subject to

$$\sum_{m=1}^{M} d_m \le d_n, \qquad (3)$$

and

$$\rho_n \le g_n^m \le R^m - \sum_{i=1}^{n-1} g_i^m, \forall m = 1, \cdots, M.$$

From Eqs. (2) and (3), we have

$$\sum_{m=1}^{M} d^m = \sum_{m=1}^{M} \left(\frac{L}{g_n^m} + \frac{L}{R^m}\right) \le d_n.$$

Owing to $\sum_{m=1}^{M} (R^m - \sum_{i=1}^{n-1} g_i^m)$ be a fixed value (because there have been n-1 connections in the route), we can transfer the allocation problem to minimize the value of $\sum_{m=1}^{M} g_n^m$.

To obtain the maximum of total residual bandwidth, thereby minimizing the sum of the bandwidth that allocated to this connection at each switching node, we first show the following lemma.

Lemma 1 *Subject to*

$$Q = \sum_{m=1}^{M} \frac{1}{g_m},$$

where Q is a constant. If $g_1 = g_2 = \cdots = g_M$ then $\sum_{m=1}^{M} g_m$ has minimum.

Proof:
By induction

(1) Let M=2 then $Q = \frac{1}{g_1} + \frac{1}{g_2}$.

Case 1: Without loss of generality (WLOG), we let $g_1 < g_2$.

Assume that $g_1 = cg_2, 0 < c < 1,$

then $\frac{1}{cg_2} + \frac{1}{g_2} = Q,$

and $g_2 = \frac{1+c}{c} * \frac{1}{Q}, \; g_1 = \frac{1+c}{Q}.$

Finally, we have

$$g_1 + g_2 = \frac{c^2 + 2c + 1}{c} * \frac{1}{Q}. \qquad (4)$$

Case 2: Let $g_1 = g_2,$

then $\frac{2}{g_1} = Q$ and $g_1 = \frac{2}{Q} = g_2,$

therefore, we have

$$g_1 + g_2 = \frac{4}{Q}, \qquad (5)$$

$$(4) - (5) = \frac{c^2 - 2c + 1}{cQ} = \frac{(c-1)^2}{cQ} > 0.$$

Hence, we derive that: If $g_1 = g_2$ then $g_1 + g_2$ have the minimum.

(2) Assume M=n, and $Q = \frac{1}{g_1} + \cdots + \frac{1}{g_n}$.

Let the following statement hold: if $g_1 = g_2 = \cdots = g_n$, then $\sum_{i=1}^{n} g_i$ has minimum.

(3) Let M=n+1.

Case 1: WLOG, let $g_1 < g_2 = \cdots = g_{n+1}$.

Then we hve $Q = \frac{1}{g_1} + \cdots + \frac{1}{g_{n+1}}.$

Assume $g_1 = c_1 g_2, 0 < c_1 < 1,$

then $\frac{1}{c_1 g_2} + \frac{n}{g_2} = Q \Rightarrow \frac{1 + nc_1}{c_1 g_2} = Q.$

Hence $g_2 = \frac{1 + nc_1}{c_1 Q}, \; g_1 = \frac{1 + nc_1}{Q},$

and

$$\sum_{i=1}^{n+1} g_i = ((1 + nc_1) + \frac{n + n^2 c_1}{c_1}) * \frac{1}{Q}. \qquad (6)$$

Case 2: Let $g_1 = g_2 = \cdots < g_{n+1}$.

Assume $g_{n+1} = c_2 g_1, c_2 > 1,$

then $\frac{1}{c_2 g_1} + \frac{n}{g_1} = Q \Rightarrow \frac{1 + nc_2}{c_2 g_1} = Q,$

and $g_1 = \frac{1 + nc_2}{c_2 Q}, \; g_{n+1} = \frac{1 + nc_2}{Q}.$

We have

$$\sum_{i=1}^{n+1} g_i = ((1 + nc_2) + \frac{n + n^2 c_2}{c_2}) * \frac{1}{Q}. \qquad (7)$$

Case 3: $g_1 = g_2 = \cdots = g_{n+1} \Rightarrow \frac{n+1}{g_1} = Q \Rightarrow g_1 = \frac{n+1}{Q},$

then

$$\sum_{i=1}^{n+1} g_i = \frac{(n+1)^2}{Q} \qquad (8)$$

Hence

$$(6) - (8) = \frac{n(c_1^2 - 2c_1 + 1}{c_1 Q} = \frac{n(c_1 - 1)^2}{c_1 Q} > 0,$$

$$(7) - (8) = \frac{n(c_2^2 - 2c_2 + 1}{c_2 Q} = \frac{n(c_2 - 1)^2}{c_2 Q} > 0.$$

As $g_1 = g_2 = \cdots = g_{n+1}$, $\sum_{i=1}^{n+1} g_i$ have the minimum value.

The lemma is proved.

Next, we consider the allocation problem again. Assume that a series of M switches that have unused bandwidth $r^1, r^2, \cdots, r^M$. WLOG, let $r^1 \le r^2 \le \cdots \le r^M$. In addition, assume the input traffic of connection n is (σ_n, ρ_n)-smooth and has an end-to-end delay requirement d_n. The maximum packet size in the network is L.

Furthermore, the bandwidth allocated in each switch are $g_n^1, g_n^2, \cdots, g_n^M$. Then, we have

$$d_n \geq \sum_{m=1}^{M} \frac{L}{g_n^m} + \sum_{m=1}^{M} \frac{L}{R^m}. \quad (9)$$

Let

$$\frac{d_n - \sum_{m=1}^{M} \frac{L}{R^m}}{L} = Q,$$

then Eq. (9) can be rewritten as

$$Q \geq \sum_{m=1}^{M} \frac{1}{g_n^m}.$$

For unwasting bandwidth, let $Q = \sum_{m=1}^{M} \frac{1}{g_n^m}$. By lemma 1, if $g_n^1 = g_n^2 = \cdots = g_n^M = g$ then $\sum_{m=1}^{M} g_n^m$ is minimized. Therefore, $g = \frac{M}{Q}$ is obtained. The bandwidth, $\frac{M}{Q}$, is intially allocated to the switch which has unused bandwidth r^1. If $g \leq r^1$, then all the switches are allocated the same service rate being equal to $\frac{M}{Q}$. Otherwise, the switch which has unused bandwidth r^1 is allocated the service rate r^1. Recalculate the service rate for the other switches as follows:

$$Q = \sum_{m=1}^{M} \frac{1}{g_n^m} = \frac{1}{g_n^1} + \sum_{m=2}^{M} \frac{1}{g_n^m} = \frac{1}{r^1} + (M-1)\frac{1}{g},$$

$$\Rightarrow Q - \frac{1}{r^1} = (M-1)\frac{1}{g},$$

$$\Rightarrow g = \frac{M-1}{Q - \frac{1}{r^1}}.$$

If $\frac{M-1}{Q - \frac{1}{r^1}} \leq r^2$, then all the switches except the switch with the residual bandwidth r^1 are allocated the same service rate being $\frac{M-1}{Q - \frac{1}{r^1}}$. Otherwise, the switch which has unused bandwidth r^2 is allocated the service rate r^2. Repeat the process until all the switches are allocated. While the service rate g_n^m is computed for each node, the delay allocated to node m is derived as $d^m = \frac{L}{g_n^m} + \frac{L}{R^m}, \forall m = 1, \cdots, M$, and the service curve can be obtained. That is, if the service curve for each node m is allocated as $S^m = \sigma + \rho(t - d^m)$, the available bandwidth in the network is maximized without violating the delay bound guaranteed for the connection. We call our method as MaxMin allocation.

The algorithm is given in the following.

Input: input traffic - (σ_n, ρ_n) - smooth, delay requirement
- d_n, packet size in the connection - L, link capacity
- $R^m, \forall m = 1, \cdots, M$.

Output: the allocated service curve $S^m, \forall m = 1, \cdots, M$.

Phase 1:

1 sort the residual bandwidth in each node such that $r^1 \leq r^2 \leq \cdots \leq r^M$.

2 $g \leftarrow \frac{ML}{d_n - \sum_{m=1}^{M} \frac{L}{R^m}}$.

3 if $g \leq r^1$

4 then $g_n^m = g, \forall m = 1, \cdots, M$

5 go to Phase 3.

6 if $g > r^1$

7 then i=1.

8 go to Phase 2.

Phase 2:

1 $g_n^i \leftarrow r^i$.

2 $Q \leftarrow Q - \frac{1}{r^i}$.

3 if i=M

4 then go to Phase 3.

5 else $g \leftarrow \frac{M-i}{Q}$.

6 if $g \leq r^{i+1}$

7 then $g_n^m = g, \forall m = i, \cdots, M$

8 go to Phase 3.

9 if $g > r^{i+1}$

10 then i=i+1.

11 go to Phase 2.

Phase 3:

1 $d^m \leftarrow \frac{L}{g_n^m} + \frac{L}{R^m}, \forall m = 1, \cdots, M$.

2 $S^m = \sigma + \rho(t - d^m), \forall m = 1, \cdots, M$.

4 Simulation

We performed simulations with two topologies. The one is tandem model, the other is cross traffic model. The simulation tool is Ptolemy, a simulator developed by Berkeley University. We compared MaxMin policy with Even division policy [7] under SCED scheduling. The Even policy allocated equal shares of the end-to-end delay among links on a path.

4.1 Tandem model

Figure 2 is the topology for our first simulation. The path is from S to D. The bandwidth between each node is 400 units except for link 3 whose bandwidth is 380 units. We created 50 connections in this path. Figure 3 shows relation between the blocking probability over the end-to-end delay requirement in the five hops tandem model. We observe that the MaxMin performs significantly better than Even.

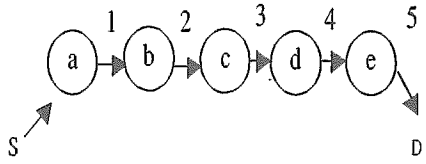

Figure 2 Tandem model

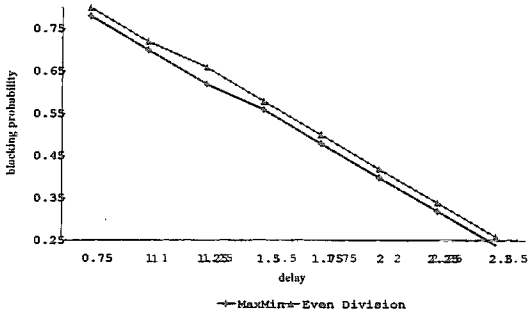

Figure 3 Blocking probability over end-to-end delay requirement in tandem model

4.2 Cross traffic model

Figure 4 is the network topology for our simulations in this section. There are three paths in this topology, S1 to D1, S2 to D2, and S3 to D3. First, we let the bandwidth for all links be 150 units except for link 4, whose bandwidth is 300 units. Every sender created 40 connections to its destination. Figure 5 shows relation between the blocking probability over the end-to-end delay requirement in cross traffic model. We observe that no matter how many delay requirements are, the blocking probability of MaxMin policy is lower than Even division policy. The second, we ranged the bandwidth of link 4 (the bottleneck) from 250 to 300 units. Figure 6 shows the results. We observe that when the bottleneck bandwidth is extended, the blocking probability is reduced. We also observe that even in cross traffic model, the MaxMin policy still can promote the network utilization efficiently.

5 Conclusions

This study presents a novel approach called MaxMin allocation policy to maximize the network's efficiency, referred as the available bandwidth, without violating each end-user's delay requirement. The proposed scheme and algorithm appropriately make the network to obtain the maximum of the network's available bandwidth.

Our proposed method is based on service curves. For promoting network utilization, our scheme allocating service curve at the servers in a tandem network such that a prespecified network service curve is guaranteed and the excess bandwidth is maximized. We reduce the service curve allocation to end-to-end delay allocation, and using MaxMin delay allocation policy, to promote the network utilization. From simulations, we find that the performance of MaxMin policy is better than Even division policy.

The concept of the excess bandwidth is to treat the resources of the network being all equal important. We can join the weighted concept to the excess bandwidth to emphasize the effect of bottleneck link in the networks. With this modification, we believe that this performance index will be more suitable to evaluate the allocation policy.

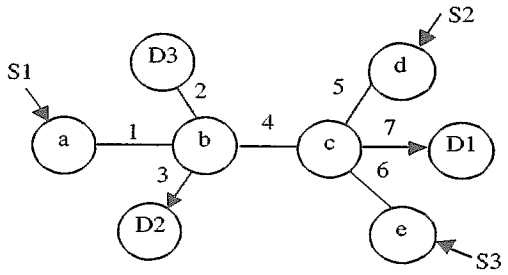

Figure 4 Cross traffic model

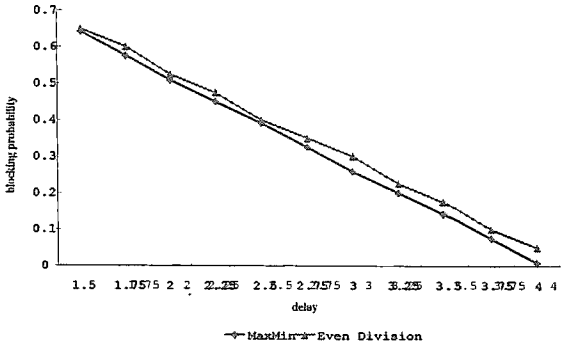

Figure 5 Blocking probability over end-to-end delay requirement in cross traffic model

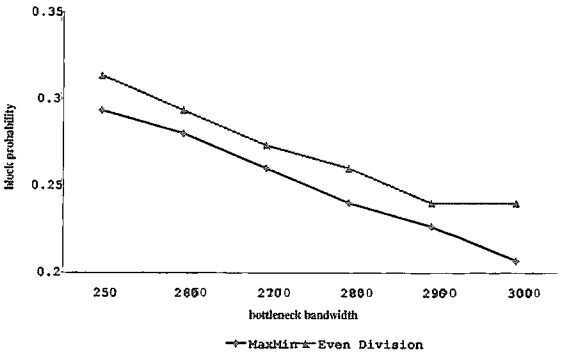

Figure 6 Blocking probability for MaxMin and Even as a Function of Bottleneck Link Capacity in cross traffic model

Reference

[1]D. Ferrari and D. Verma ,"A Scheme for Real-time Channel Establishment in Wide-Area Network",IEEE Journal Selected Areas of communications,8(4),pp.368-379,April 1990.

[2]Edward W. Knightly, and Paola Rossaro,"Improving QoS

through Traffic Smoothing",Proceedings of IFIP IWQoS'96,Paris,France.
[3]H. Zhang and D. Ferrari, "Rate-controlled service disciplines.", Journal of High Speed Networks, Vol.3, No.4, pp.389-412, 1994.
[4]H. Sariowan, Rene L. Cruz, "Scheduling for Quality of Service Guarantees via service curve",Proceeding of ICCCN, 1995.
[5]Rene L. Cruz, "Quality of Service Guarantees in Virtual Cirvuit Switched Networks",IEEE Jouranl on Selectd areas in Communications, 1995.
[6]Rene L. Cruz, George C. Polyzos, Walter Burkhard, Elias Masry,and Ramesh Rao, "A Service-Curve Approach to Performance Guaranrees in Integrated-Service Networks ",Ph.D. Dissertation in Electrical and Computer Engineering, University of California, San Diego,1996.
[7]Victor Firoiu and Don Towsley, "Call Admission and Resource Reservation for Multicast Sessions", Proceedings of IEEE INFOCOM'96.
[8]Y. P. Chu, E. H. Hwang, K. C. Lin and C. H. Chen,"Local Allocation of End-to-end Delay Requirement",IEICE Trans. on Communication,1999.
[9]Y. P. Chu, E. H. Hwang, K. C. Lin and C. H. Chen, "Efficient Allocation For Quolity of Service Guarantees",Technical Reprot,1998.
[10]Y. P. Chu, E. H. Hwang, K. C. Lin and C. H. Chen, "MaxMin Allocation -A Benchmark For QoS Guarantees Schemes",Technical Reprot,1998.