On Optimal End-to-End QoS Budget Partitioning in Network Dimensioning*

Hyunjoon Cho[1], André Girard[2], and Catherine Rosenberg[3]

[1] School of Electrical and Computer Engineering, Purdue University, hyunjoon@purdue.edu
[2] INRS-EMT, Université du Québec, andre@emt.inrs.ca
[3] Department of Electrical and Computer Engineering, University of Waterloo, cath@ece.uwaterloo.ca

Abstract: We investigate the problem of optimal end-to-end QoS budget partitioning to quantify the advantage for network dimensioning of having a non-uniform allocation of end-to-end QoS requirement over the links in a path. We extend a previous revenue maximization model to M/G/1 queuing models and propose a fast partitioning heuristic based on the M/M/1 case. We then show on small networks with M/G/1 queues that the heuristic gives a near-optimal partitioning and confirm previous numerical results obtained for the M/M/1 model that optimal partitioning can bring large cost reductions as compared with equal partitioning.

Keywords: QoS budget partitioning, network dimensioning, effective bandwidth

1. INTRODUCTION

The current Internet runs with best effort and does not support any explicit performance guarantees. However, the demand for guaranteed services has been growing rapidly due to the advent of real-time applications such as multimedia streaming and voice over IP. In this context, where many sessions may have a set of end-to-end Quality of Service (QoS) requirements, some network resources must be managed on an end-to-end basis to guarantee the necessary performance. It follows that some form of admission control is needed so that a request can be rejected if accepting it would make the network unable to guarantee the QoS constraints for existing sessions.

Computing the end-to-end QoS bounds for all affected sessions on the new session's route at each connection setup time is cumbersome and time-consuming if the network has a large number of different o-d pairs. This problem can be solved by partitioning the end-to-end QoS requirement into individual link QoS assignments. A new connection can then be admitted only when all the link QoS guarantees on a path are satisfied and this can be checked locally. The end-to-end QoS budget partitioning problem is to find the QoS link allocation that is best according to some relevant objective function.

*This work was supported in part by an NSF grant (contract no. 0099137-ANI) and an NSERC grant (contract no. 121949-00).

The motivation for local allocation approach was discussed extensively in [1]. There has also been a large amount of work on the optimal partitioning of the end-to-end QoS budget into local constraints along a path [2, 3, 4, 5]. The work in this field assumed that a good allocation scheme is in fact needed and proposed methods to do this efficiently. Designing QoS allocation algorithms, on the other hand, is useful only when optimal partitioning brings any substantial gain in network performance. The question of what can be gained from the optimal or near-optimal allocations was first examined in [1]. The authors found that the relative performance of partitioning policies is heavily dependent on the QoS metric. The work of Diwan et al. [6], based on simulations with uniform traffic, concluded that there is no advantage in optimal per-node rate allocation to provide end-to-end delay guarantees. Comparison of equal and optimal partitioning with envelope-regulated traffic for a tandem network was discussed in [7]. These results on the gains of optimal partitioning are inconclusive since they rely on different QoS measures and the evaluation is done under different conditions based on particular traffic models, network structures, and cost functions.

This has been the motivation to provide in [8] a unified framework to quantify the advantage of having a non-uniform allocation of the end-to-end QoS budget over the links in a path. Preliminary results have indicated that this framework can yield useful insights on the value of optimal allocation and that this optimal allocation can bring significant savings. One of its main features is that using a variety of different traffic sources is extremely simple once an effective bandwidth is available. The first result of the present paper is to illustrate this point by using a different traffic model and pointing out precisely where changes have to be made to the design problem.

We also present an interesting decomposition property of the dimensioning problem when the packet queues are M/M/1. In that case, we find that the optimal delay budget partitioning can be done *independently* of traffic matrices and routing. We discuss the benefits of this decomposition property and show that it is quite accurate for some queuing models other than M/M/1. Finally, we present numerical results showing that the results of [8] on the savings available from optimal partitioning are still valid with a different queueing model.

The paper is organized as follows. In Section 2, we present a short summary of the dimensioning problem that has been described in detail in [8]. In Section 3, we show how easily we can handle various traffic models. Some numerical results are then shown in Section 4. We solve the nonlinear program and show the advantage of the optimal partitioning over the equal partitioning with various effective bandwidth models in network dimensioning. Finally, we end in Section 5 with some concluding remarks and future directions.

2. PROBLEM FORMULATION

First we summarize some of the key concepts used in [8]. Because the QoS requirements are end-to-end, we assume that the network protocols support the concept of end-to-end connection or session. This is not unrealistic in the context of MPLS or with the notion of packet flows. We also assume that the network can reject a request for connection if the current conditions are not adequate. A third assumption is that the packet QoS performance requirements of a session can be guaranteed by giving it sufficient bandwidth at each link along the path the session occupies in the network. This assumption leads to the notion of *effective bandwidth* (EB) [9, 10, 11]. Routing is by load sharing with coefficients $\alpha_i^{l,m}$ which indicate the fraction of traffic demand between o-d pair (l, m) is offered to the i^{th} path between these nodes. There is no retrial in case of blocking. With these assumptions, the network design problem of [8] is an extension of the revenue maximization model defined in [12]. In this paper, we denote paths through the network by a triplet (l, m, i) where l denotes the origin of the path, m its destination, and i an index to represent the i^{th} path in a list of paths from l to m.

2.1. Delay Partitioning

The problem of QoS budget allocation is quite general and applies to different QoS metrics, be it average delay, jitter, packet loss, etc. In order to simplify the discussion, we assume here that the QoS budget that is to be partitioned is the average packet delay through the network. Much of the modeling and analysis applies equally well to other measures of QoS as long as they are approximately additive on a path.

The QoS constraint is expressed as a bound $\overline{D}^{l,m}$ on the average delay experienced by packets on the path from their origin l to their destination m, the so-called end-to-end delay budget. The question is then how to choose the local delay bound $d_s^{l,m}$ for each o-d pair (l, m) on each link s given the value of $\overline{D}^{l,m}$.

A straightforward solution is to define a network-wide reference path of maximum length K and to allocate a local delay constraint $d_s^{l,m,i} = \overline{D}^{l,m}/K$ for all flows on the i^{th} path for o-d pair (l, m) that use link s. This we call *Reference Partitioning* (RP) and it is the technique that is usually used in current networks. A more sophisticated technique would be to allocate the delay budget equally for *each* path separately, i.e., set $d_s^{l,m,i} = \overline{D}^{l,m}/K_i^{l,m}$ for all s in the path where $K_i^{l,m}$ is the number of links on path (l, m, i). This we call *Equal Partitioning* (EP). Although these two techniques have the obvious merit of simplicity, we can ask how more efficient it would be to partition the delay budget in the best possible way. This we call *Optimal Partitioning* (OP).

Here we consider only a FIFO queueing discipline so that all the connections on a link experience the same average delay. The link delay bound on a link can then be written

$$d_s = \min_{(l,m,i)} \{ d_s^{l,m,i} \mid s \in (l, m, i) \}. \tag{1}$$

The problem is then to determine d_s for each link s to minimize some suitable objective function along with other relevant design parameters.

2.2. Optimization Model

In this paper, we use the term *Quality of Service* to describe the performance parameters for *packets*. Because we are using an Effective Bandwidth technique, the only real-time decision that has to be made is whether to accept or reject a connection request since the EB model guarantees that all the QoS constraints will be automatically met if there is enough bandwidth to accept it. The measure of performance for a network is then the probability of rejection of a request, denoted $L^{l,m}$ for o-d pair (l, m). This we call the *Grade of Service* (GoS) to stress the fact that it applies to connections instead of packets.

We can state the optimization problem for the combined calculation of the routing $\boldsymbol{\alpha}$, delay allocation $\mathbf{d}$, and dimensioning $\mathbf{C}$ with a capacity cost function $g_s(C_s)$ for link s. We get

$$\min_{\boldsymbol{\alpha},\mathbf{C},\mathbf{d},\mathbf{B}} \mathcal{V}(\boldsymbol{\alpha},\mathbf{C},\mathbf{d},\mathbf{B}) = \sum_s g_s(C_s) \tag{2}$$

$$\sum_i \alpha_i^{l,m} = 1, \quad \alpha_i^{l,m} \geq 0, \tag{3}$$

$$L^{l,m} \leq \overline{L}^{l,m}, \tag{4}$$

$$\sum_s d_s I_{s,(l,m,i)} \leq \overline{D}^{l,m}, \quad d_s \geq 0, \tag{5}$$

$$E[a_s, N_s(C_s, d_s)] = B_s, \tag{6}$$

where $I_{s,(l,m,i)}$ is the indicator function such as $I_{s,(l,m,i)} = 1$ when link s is in path (l,m,i). Note that the artificial independent variables B_s, representing the link blocking probability, and the corresponding constraints (6) are added to avoid the difficulty of the fixed-point system in the calculation. The Erlang B function is denoted by $E(a_s, N_s)$ where a_s is the connection arrival rate on the link and N_s is the number of servers. N_s is calculated as the maximum number of sessions that can be present on the link while still meeting the packet QoS requirements, and it is a function of C_s and d_s. More details can be found in [8, 13].

If we use RP or EP, the variables $\mathbf{d}$ are fixed. In that case, the N_s in Eq. (6) is a function of C_s only since the d_s's are now given and the optimization is performed with respect to $\boldsymbol{\alpha}, \mathbf{B}$, and $\mathbf{C}$ only.

3. MODEL APPLICATION

The formulation (2–6) is a simplification to the case of network dimensioning of the general model of [8]. In this section, we show how easily the model can be applied to different queueing models and discuss an interesting property of the M/M/1 EB model.

3.1. Effective Bandwidth Model for M/G/1 Queue

One of the main advantages of the formulation of (2–6) is that different traffic sources can be introduced very simply in the dimensioning procedure once an effective bandwidth is known. This is done here with the M/G/1 delay model. Note that we *do not* claim that the M/G/1 model is a realistic description of packet flows in real networks. It is chosen because it is analytically simple and there is a known EB model for it. For the M/G/1 queue, the Effective Bandwidth W of a session subject to a bound w on the average packet waiting time before service was proposed by Kelly in [10] as

$$W(w) = \lambda \left[\mu + \frac{1}{2w}\left(\mu^2 + \sigma^2\right)\right], \tag{7}$$

where packets are generated by a Poisson process with rate λ and the service time is arbitrary with mean μ and variance σ^2. Assuming a link has capacity C, the average packet length is $\overline{p}$, and the variance of the service time $\sigma^2 = \beta\left(\overline{p}/C\right)^2, \beta \geq 0$, the maximum number of connections N on a link can be computed as

$$N(C,d) = \frac{2C(dC - \overline{p})}{\lambda\overline{p}[(\beta - 1)\overline{p} + 2dC]} \tag{8}$$

where d is the average packet delay including the service time as the QoS constraint. This equation is used to compute the number of servers in the Erlang B function and we write $N(C, d)$ to emphasize the fact that it is a function of the C and d variables which are the decision variables of our problem. This shows one of the main advantages of our formulation in that the model is particularly easy to use with different types of packet traffic since this is the *only* point in the model where the parameters of the packet process appear. Setting $\beta = 0$ in Eq. (8) yields the value of N for an M/D/1 queue and $\beta = 1$ for M/M/1.

3.2. Delay Allocation Heuristic

We now discuss the decomposition property of the M/M/1 EB model and its potential advantage for network dimensioning. For the M/M/1 queue, the number of servers in a link can be computed from Eq. (8) by setting $\beta = 1$ as

$$N = \frac{1}{\lambda}\left(\frac{C}{\overline{p}} - \frac{1}{d}\right). \tag{9}$$

Recall that the objective of the optimization problem is to minimize the capacity construction cost while satisfying the GoS and QoS constraints. From Eq. (9), we can write

$$C_s = \overline{p}\left(\lambda N_s + \frac{1}{d_s}\right). \tag{10}$$

If we assume the construction cost function is linear, i.e. $g_s(C_s) = \gamma_s C_s$ for some constant γ_s, the objective function can be rewritten as

$$\mathcal{V} = \sum_s \gamma_s C_s = \overline{p}\left(\lambda \sum_s \gamma_s N_s + \sum_s \frac{\gamma_s}{d_s}\right). \tag{11}$$

The objective thus separates into two parts: one for the $\mathbf{N}$ and $\boldsymbol{\alpha}$ and the other for the $\mathbf{d}$. Given that the set of constraints also separates into two independent sets, (5) for the $\mathbf{d}$ variables and the others for the $\mathbf{N}$ and $\boldsymbol{\alpha}$, the overall dimensioning problem separates into two *independent* subproblems. Then, the optimal delay allocation is simply the solution of

$$\min_{\mathbf{d}} \; \mathcal{V}_d = \sum_s \frac{\gamma_s}{d_s},$$
$$\sum_s d_s I_{s,(l,m,i)} \leq \overline{D}^{l,m}, \quad d_s \geq 0,$$

and is completely independent of the traffic since it depends only on the topology of the network and on the paths that are used to carry the traffic. On the other hand, finding the optimal $\mathbf{N}^*$ is the classical network dimensioning problem of circuit-switched networks which can be solved by known algorithms.

The decomposition property is valid for any queueing model where the expression for $C_s(N_s, d_s)$ obtained from the maximum number of connections such as Eq. (8) separates into two independent functions of the form $C_s = h_1(N_s) + h_2(d_s)$. This is rather unlikely to happen other than for the M/M/1 case, and is not true even for M/D/1. Nevertheless, its simplicity makes it an attractive heuristic for the computation of the delay allocation when the EB formula does not separate. In what follows, the partitioning derived from the M/M/1 EB model will be called *Heuristic Partitioning* (HP). The question of course is how accurate this approximation would be in these cases. This question will be examined numerically in Section 4.

4. NUMERICAL RESULTS

In this section, we extend the results of [8] to quantify the advantage of optimal partitioning in network dimensioning. The numerical results presented here are based on the M/M/1, M/D/1, or M/G/1 queuing model with average delay as the sole QoS requirement. They are presented to show the flexibility afforded by the formulation of [8]. We limit the exposition to very small networks in order to be able to understand clearly the solutions we get. We feel that this is sufficient to show that the model can provide useful results and that significant savings are possible. The networks that we use for the numerical examples do not involve any routing decision so that the $\boldsymbol{\alpha}$ variables can be ignored.

4.1. M/M/1 Effective Bandwidth

First we solve the model (2–6) for Poisson traffic and reproduce the results obtained in [8] with the M/M/1 EB model. We use a simple concentration network with a single backbone link and several edge links as shown in Figure 1. Although the network is of a very specific structure, it is a form frequently observed in real networks [14].

Results are presented for the relative gain G obtained from the optimal partitioning over the equal partitioning, expressed as $G = (\mathcal{V}^E - \mathcal{V}^O)/\mathcal{V}^E$, where $\mathcal{V}$ is the objective function of (2) and the superscripts E and O denote the equal and the optimal partitioning respectively. We found that the average packet length $\overline{p}$ does not affect the relative gain. Hence, we show the results varying the following 4 parameters: the number n of edge nodes, $\lambda, A^{l,m}$, and $\overline{D}^{l,m}$. In all figures, the relative gain is plotted versus the end-to-end delay constraint while varying the other parameters. We assume there are n o-d pairs in Figure 1 and node l has l' as its destination, which makes a path P_l. In order to simplify the presentation, we use $\overline{L}^{l,l'} = \overline{L}$, $\overline{D}^{l,l'} = \overline{D}$, and $A^{l,l'} = A_0$ for all (l, l') pairs. We also assume $g_s(C_s) = C_s$ for all s.

Figure 2 shows the relative gain as a function of the delay constraint for various values of n. We see from the figure that the gain gets larger as n increases and can go up to 20% but is much smaller if n is small, e.g., $n = 5$, where the gain is less than 10%.

We can see how the actual solutions differ between EP and OP in Table 1 where the

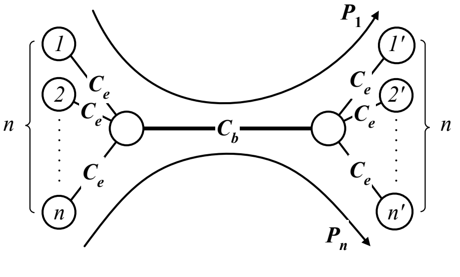


Figure 1: Sample network.

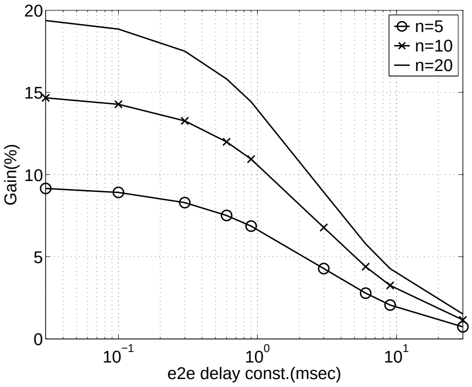


Figure 2: Gain vs number of edges for $\lambda = 50, A_0 = 10$, and $\overline{L} = 0.01$.

Table 1: Dimensioning results for the concentration network, $\overline{D} = 1$ ms. Capacities are in unit 10^3 and delays in ms.

	$n = 5$				$n = 10$				$n = 20$			
	d_b	d_e	C_b	C_e	d_b	d_e	C_b	C_e	d_b	d_e	C_b	C_e
EP	0.33	0.33	6.45	3.93	0.33	0.33	9.31	3.93	0.33	0.33	14.8	3.93
OP	0.18	0.41	8.92	3.38	0.14	0.43	13.6	3.25	0.10	0.45	21.7	3.15

indices e and b denote the edge and backbone links respectively. There is a substantial difference both in the delay and the capacity between the two solutions. The delay on the backbone is reduced and the capacity increased in the OP solution as compared to EP while the opposite is true for the edge links. In other words, we get a lower cost if we tighten the bound and increase the capacity in the core and relax it in the access.

Figure 3 shows the relative gain for $n = 20$ with several λ's, and Figure 4 for different connection arrival rates. The impact of the two parameters on the gain is similar: we get a larger gain with a smaller packet or connection arrival rate and the spread is larger at intermediate values of the QoS bound.

We can conclude from these results that the parameters of the system can have an

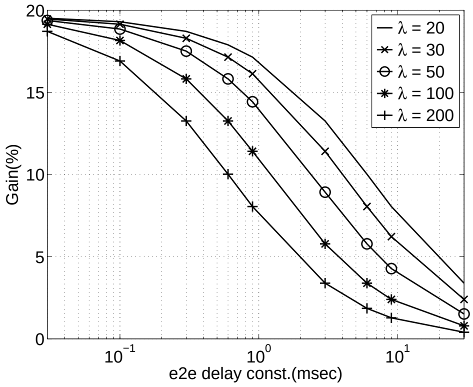


Figure 3: Gain vs packet arrival rate for $A_0 = 10$, and $\overline{L} = 0.01$.

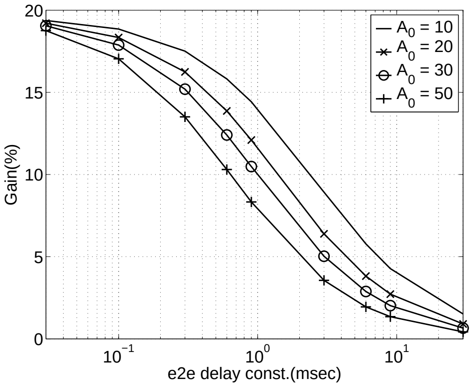


Figure 4: Gain vs connection request arrival rate for $\lambda = 50$, and $\overline{L} = 0.01$.

effect on the gain that can be obtained from optimal partitioning in network dimensioning. For a given delay bound, the most important parameter seems to be n and, indirectly, the distribution of traffic in the network since when n is large, the amount of traffic on the backbone link is very much larger than that of an edge link. When each o-d pair has a large connection arrival rate or each connection has a large packet arrival rate, we could get a significant gain only when the delay constraint is very tight.

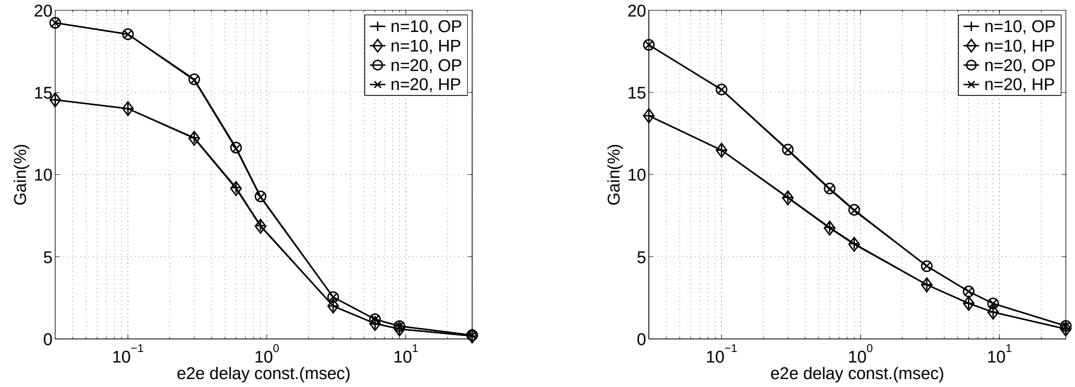
4.2. M/D/1 and M/G/1 Queues

We also solve the same dimensioning problem for the M/D/1 and M/G/1 queueing models. We want to see if the conclusions obtained in Section 4.1 still hold and check how the HP heuristic performs with non-M/M/1 models by comparing it with the actual optimal solution. OP is obtained by solving the complete optimization problem with $\mathbf{C}$ and $\mathbf{d}$ as decision variables. HP, on the other hand, is obtained by minimizing $\sum_s 1/d_s$ with the constraints (5), and then, the optimization is done with the remaining variables by substituting the d_s's in Eq. (8).

The gains of OP and HP over EP are shown in Figure 5 as a function of the delay constraint with the M/D/1 and an M/G/1 EB models. The construction cost with HP are almost the same as those obtained through the complete optimization. There is little difference between the gains of OP and HP over EP, though the curves of the M/G/1 model are a little flattened out compared with the M/D/1 case.

We plotted in Figure 6 the gains of OP and HP with several β's. Note that the end-to-end delay constraint of Figure 6.b is 10 times smaller than that of Figure 6.a. With a relatively loose delay bound, allocating local delay constraints with HP may end up with a slightly higher cost, and thus lower gain, than with the complete optimization, especially when β has a large value. However, if the delay bound is very tight, there is no significant cost increase caused by using HP even when the packet length distribution is very different from exponential. This can be explained from the expression of the capacity given by Eq. (8)

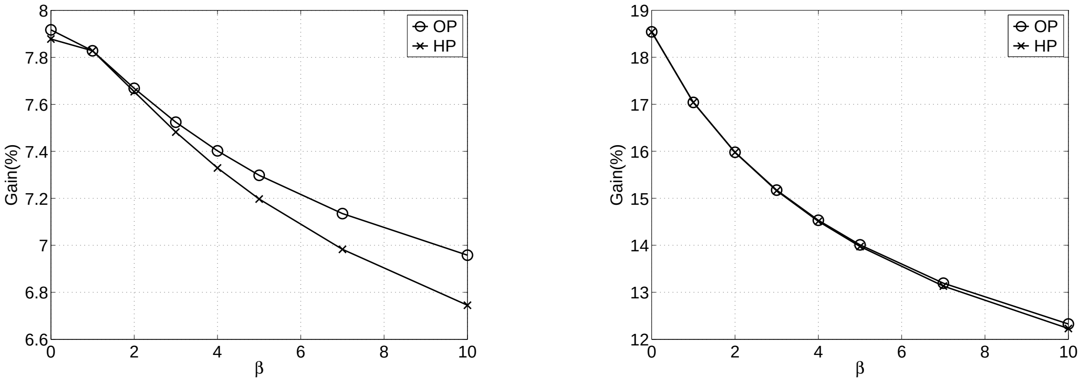
$$C = \frac{\lambda \overline{p} N}{2} + \frac{\overline{p}}{2}\left[\frac{1}{d} + \sqrt{(\lambda N)^2 + \frac{2\beta\lambda N}{d} + \frac{1}{d^2}}\right]. \tag{12}$$



a. M/D/1

b. M/G/1 ($\beta = 3$)

Figure 5: Gain vs $\overline{D}$ of OP and HP over EP for $\lambda = 50, A_0 = 50$ and $\overline{L} = 0.01$.



a. $\overline{D} = 1\text{ms}$ b. $\overline{D} = 0.1\text{ms}$

Figure 6: Gain vs β of OP and HP for $\lambda = 50, A_0 = 50$ and $\overline{L} = 0.01$.

If d tends toward 0, the dominant term under the square root would be the $1/d^2$ term assuming $\lambda N = o(d^{-1})$. This gives an approximate expression

$$C \approx \overline{p}\left(\frac{\lambda N}{2} + \frac{1}{d}\right) \quad \text{as} \;\; d \to 0, \tag{13}$$

which is quite a similar expression to Eq. (10) of the M/M/1 case. Hence, with a very tight delay constraint, HP would be almost the same as OP and as the result, it makes little difference in the capacity. When the delay is not small, the approximation gets closer to the optimal when β has a smaller value, as can be seen from Figure 6.

The computation results with two different connection arrival rates are shown in Table 2 for $n = 20, \lambda = 50, \overline{D} = 1\text{ms}$, and $\beta = 5$. The difference of delay allocations between OP and HP when $A_0 = 100$ is slightly larger than that when $A_0 = 10$. However, there is little difference of total construction costs between OP and HP.

5. CONCLUSIONS

We have extended the results of [8] on the optimal partitioning of the end-to-end QoS constraints. First, we have shown with an M/G/1 queuing model the flexibility afforded by this framework to compare different traffic sources. Using the M/M/1, M/D/1, and M/G/1 EB models with the average delay as a measure of QoS and small networks, we have provided quantitative results on the benefit of the optimal partitioning over an equal

Table 2: Computation results with various connection arrival rates. Capacities are in unit 10^3 and delays in ms.

A_0	Policy	$\sum_s C_s$	d_e	d_b	C_e	C_b	B_e	B_b	N_e	N_b
10	EP	220.2	.3333	.3333	5.035	18.77	.004756	.0005143	18.57	238.8
	OP	195.6	.4453	.1093	4.175	28.61	.004685	.0006585	18.59	237.5
	HP	195.7	.4497	.1006	4.150	28.61	.004679	.0006716	18.59	237.4
100	EP	593.7	.3333	.3333	12.03	112.5	.004614	.0008007	120.5	2080.
	OP	559.1	.4418	.1163	10.81	126.7	.004589	.0008516	120.5	2079.
	HP	559.6	.4497	.1006	10.74	129.8	.004585	.0008578	120.5	2079.

partitioning policy. The results confirmed the conclusions previously obtained for the M/M/1 case and showed that optimal partitioning can reduce the capacity construction cost of networks by up to 20% depending on the value of the QoS constraint. In the case of the M/M/1 EB model, we observed that the delay allocation can be done independently of the traffic intensity. This has suggested a heuristic rule for delay allocation. We have verified that this approximate rule is quite accurate for queueing models such as M/D/1 and M/G/1.

The numerical work in this paper has been done with rather simple networks and some simplified QoS models. An extensive examination with general and more complex networks and traffic is currently being performed.

REFERENCES

[1] R. Nagarajan, J. Kurose, and D. Towsley. Local allocation of end-to-end quality-of-service in high-speed networks. In *Proc. of IFIP Workshop on Performance Analysis of ATM Systems*, pages 99–118, January 1993.

[2] R. A. Guérin and A. Orda. QoS routing in networks with inaccurate information: Theory and algorithms. *IEEE/ACM Trans. on Networking*, 7(3):350–364, 1999.

[3] D. H. Lorenz and A. Orda. QoS routing in networks with uncertain parameters. *IEEE/ACM Trans. on Networking*, 6(6):768–778, 1998.

[4] D.H. Lorenz and A. Orda. Optimal partition of QoS requirements on unicast paths and multicast trees. *IEEE/ACM Trans. on Networking*, 10(2):102–114, 2002.

[5] A. Orda and A. Sprintson. A scalable approach to the partition of QoS requirements in unicast and multicast. In *Proc. of IEEE INFOCOM*, volume 2, pages 685–694, June 2002.

[6] A. Diwan, J. Kuri, and A. Kumar. Optimal per-node rate allocation to provide per-flow end-to-end delay guarantees in a network of routers supporting guaranteed service class. In *Proc. of IEEE ICC*, volume 2, pages 1112–1117, April 2002.

[7] A. Girard, C. Rosenberg, and H. Cho. Optimal performance partitioning for networks with envelope-regulated traffic. In *Proc. of ITC Specialist Seminar*, July 2002.

[8] H. Cho, A. Girard, and C. Rosenberg. On the advantages of optimal end-to-end QoS budget partitioning. Submitted to Telecommunication Systems, January 2005.

[9] F. C. Harmantzis, D. Hatzinakos, and I. Lambadaris. Effective bandwidths and tail probabilities for gaussian and stable self-similar traffic. In *Proc. of IEEE ICC*, volume 3, pages 1515–1520, May 2003.

[10] F.P. Kelly. Effective bandwidths at multi-class queues. *Queueing Systems*, 9:5–16, 1991.

[11] W. Whitt. Tail probabilities with statistical multiplexing and effective bandwidths in multi-class queues. *Telecommunication Systems*, 2:71–107, 1993.

[12] A. Girard. Revenue optimization of telecommunication networks. *IEEE Trans. on Communications*, 41(4):583–591, 1993.

[13] A. Girard. *Routing and Dimensioning in Circuit-Switched Networks*. Addison-Wesley, 1990.

[14] A.-L. Barabasi, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272:173–187, 1999.