

# TRAFFIC SAVINGS THROUGH A CAC PROCEDURE INCORPORATING A FAIR BANDWIDTH ALLOCATION POLICY FOR ELASTIC SERVICES

We apply the max-min fairness policy (MMF) for elastic services with min and max bandwidth requirements (approximated MMF) not to a single link but to a whole connection-oriented network. To achieve it we propose a Call Admission Control (CAC) procedure consisting of several scenarios dependent on the number of paths between the origin-destination (O-D) network nodes. We evaluate the global network performance at the call level by comparing the maximum Call Blocking Probability (CBP) of the proposed scheme with that of a no fair bandwidth allocation policy scheme. Simulation results show significant traffic savings in the case of the approximated MMF policy.

**Keywords:** bandwidth allocation, max-min fairness, elastic traffic, call blocking probability.

## 1. Introduction

Elastic services keep the call-level study in connection-oriented networks an open issue. We examine the applicability of a fair bandwidth allocation (FBA) policy to a whole connection-oriented network. From the bandwidth management point of view, a connection-oriented network is a mesh network. We consider connection-oriented (VC routing) networks, in which each path connecting an O-D pair has a certain bandwidth capacity (e.g. Virtual Path in ATM networks). We assume that the O-D path bandwidth is determined at the network design phase (NDP) and used only by calls of this O-D pair. Several FBA policies based on the MMF policy [1] have been proposed as a flow/congestion control mechanism of elastic traffic and tested either in a single link or in an isolated path. In a network most studies apply the MMF, either in the domain of flow/congestion control, dealing with low-level traffic [2] or in the routing domain [3] dealing with call-level traffic. We adopt the MMF policy with min and max bandwidth requirements (BR) (we call it approximated MMF, a-MMF) as the most appropriate for elastic traffic and use a linear programming (LP) model to present it [4]. We aim at applying the a-MMF to the call-level of a network accommodating elastic service-classes while we concentrate on the CAC process, where bandwidth is assigned to each call. We propose the application of the a-MMF at call setup among calls of the same destination taking into account the available bandwidth, *avl* of the O-D path while avoiding the bandwidth modification of in-service calls which requires a feedback control mechanism for all service-classes. To achieve our purpose we devote a fraction of the max permitted time for the completion of call-setup to the a-MMF process. The ultimate goal is to improve the global call-level network performance.

In section 2 we propose a LP model in order to provide a rigorous description of the a-MMF. In section 3 we present the CAC procedure that makes possible the application of the a-MMF to

the entire network. Section 4 contains an application example for evaluation. We conclude in section 5.

## 2. LP model for the approximated MMF policy

The MMF policy with min and max BR of the services cannot be presented by a single LP model. Nevertheless, we propose an approximate single LP model which has the advantage of a clear and rigorous description of the a-MMF; its solution is not always an MMF vector. If one omits the min and max BR, then the resultant bandwidth vector is an MMF vector.

Let  $S$  and  $P$  be the set of network service-classes and O-D paths, respectively. Calls are conveyed by a fixed routing scheme. By  $C_p$  we denote the *avl* of O-D path  $p \in P$ , by  $S_p$  the set of service-classes which use O-D path  $p \in P$ , by  $n_s$  the number of calls of service-class  $s \in S_p$  and by  $r_s$  the bandwidth which will be allocated to new calls of service-class  $s \in S$ . The following LP model provides a rigorous description of the a-MMF.

The constraint sets of the LP model are the following [4]:

$$\text{Network set: } \sum_{s \in S_p} n_s r_s \leq C_p, \text{ for all } p \in P \quad (1)$$

$$\text{Traffic parameters set: } b_{\min,s} \leq r_s \leq b_{\max,s}, \text{ for all } s \in S \quad (2)$$

$$\text{Policy set: } r_s \geq \min \left\{ \frac{C_p}{\sum_{s \in S_p} n_s}, b_{\max,s} \right\}, \text{ for all } s \in S_p \quad (3)$$

$$\text{The objective function is: } \sum_{s \in S} w_s r_s \Rightarrow \max \quad (4)$$

\* Michael D. Logothetis, Ioannis D. Moscholios and George K. Kokkinakis

Wire Communications Laboratory, Dept. of Electrical & Computer Engineering, University of Patras, Greece. E-mail: m-logo@wcl.ee.upatras.gr

where  $b_{min,s}$ ,  $b_{max,s}$  are the min, max BR per call of  $s \in S$ , respectively, and  $w_s$  is a weight (priority) associated with each  $s \in S$  (if no priority is defined, then  $w_s = 1.0$  for each  $s \in S$ ).

To solve the above LP model, we use either the primal cutting-plane algorithm (an integer programming model based on the Simplex [5]) when the bandwidth is quantized or the Simplex when the bandwidth is not quantized.

*Numerical Example:* Consider a single link of capacity 16 and four service-classes A - D, with weights (priorities)  $w_A = 2.5$ ,  $w_B = 4$ ,  $w_C = 0.5$  and  $w_D = 1$ . The number of calls arriving to the link and the corresponding bandwidth bounds of each service-class are:  $n_A = 1$  with bounds  $0 \leq r_A \leq 4$ ,  $n_B = 1$  with  $0 \leq r_B \leq 2$ ,  $n_C = 1$  with  $0 \leq r_C \leq 10$ , and  $n_D = 1$  with  $0 \leq r_D \leq 4$ . Then, the LP Model describing the a-MMF is:

*Constraints:*

(Network Set)  $r_A + r_B + r_C + r_D \leq 16$   
 (Traffic parameters set)  $0 \leq r_A \leq 4$ ,  $0 \leq r_B \leq 2$ ,  $0 \leq r_C \leq 10$ ,  
 $0 \leq r_D \leq 4$

(Policy Set)  $r_A \geq \min\{16/(1+1+1+1), 4\}$ ,  
 $r_B \geq \min\{16/(1+1+1+1), 2\}$ ,  
 $r_C \geq \min\{16/(1+1+1+1), 10\}$ ,  
 $r_D \geq \min\{16/(1+1+1+1), 4\}$

*Objective function:*  $2.5r_A + 4r_B + 0.5r_C + 1r_D \rightarrow \max$

Solution (by the Simplex):  $r_A = 4$ ,  $r_B = 2$ ,  $r_C = 6$ ,  $r_D = 4$ . The max-min fair share algorithm presented in [6] (page 216) produces the same results.

### 3. CAC incorporating the approximated MMF policy

To apply the a-MMF policy to a network we consider the following: i) The route and the bandwidth of the paths between each O-D pair are predetermined. ii) Calls arrive in a node according to a Poisson process and ask for call setup; no global controller exists to manage the bandwidth allocation process. iii) Each node services the calls with a CAC procedure on a FCFS basis for an exponentially distributed holding time. iv) The bandwidth of the

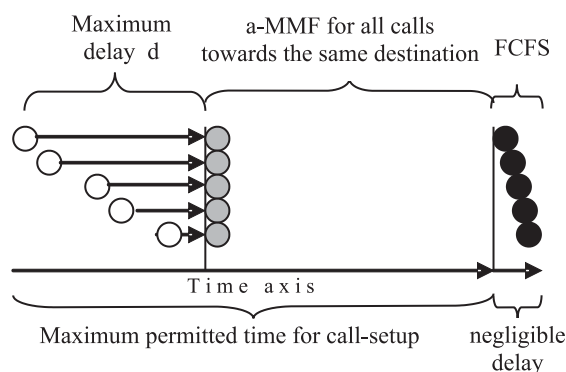


Fig. 1. The proposed new CAC scheme.

newly arriving calls must be determined and, in view of fairness, the bandwidth of in-service calls may be modified. This modification is difficult to be implemented because feedback control for all service-classes is required. We avoid this modification by proposing the following procedure: A fraction of the max permitted time of the completion of call setup is devoted to the a-MMF process (Fig. 1). CAC delays the call setup for at most  $d$  sec while waiting for new arrivals within  $d$ . During  $d$ , call releases are performed. At the end of  $d$ , new calls are grouped according to their destination; then the a-MMF policy is applied among those towards the same destination taking into account only the  $avl$  of the O-D path. Having assigned the bandwidth to each call, the call-setup process starts for all waiting calls on a FCFS basis. More precisely, the proposed procedure is as follows:

#### Approximated MMF Policy and Call Setup

For all calls towards the same destination:

- (1) **if**  $avl < \sum b_{min}$  **then:** (select scenario a, or b, or c)
  - a. All calls are blocked and lost.
  - b. Call setup is performed in the shortest path (SP) only, with  $b_{min}$ , on a FCFS basis.
  - c. Call setup is performed with  $b_{min}$ , on a FCFS basis (not only in the SP). **Endif.**
- (2) **if**  $avl \geq \sum b_{max}$  **then:**  
 Call setup is performed with the  $b_{max}$ . **Endif.**
- (3) **if**  $\sum b_{min} \leq avl < \sum b_{max}$  **then:**  
**if multiple paths exist between O-D then:**  
 Call setup is performed (on a FCFS basis):
  - In the SP with  $b_{max}$ , indicating a higher priority in the use of SP
  - In the alternative paths (AP) except the last, with  $b_{min}$ , indicating a lower priority in using them.
  - In the last path according to the a-MMF policy.**else if a single path exists between O-D then:**  
 Call setup is performed according to the a-MMF policy, on a FCFS basis. **Endif.**

### 4. Application example - evaluation

We consider the network of Fig. 2, which accommodates 3 service-classes. The 1<sup>st</sup> is the telephone service of 64 Kbps, while the 2<sup>nd</sup> and the 3<sup>rd</sup> are elastic services capable to change their bandwidth rates per call, from 128 to 384 Kbps, and from 768 to 1536 Kbps, respectively, in steps of 64 Kbps. We dimension (design) the network assuming an equal traffic-load among the switching pairs, which is 240, 40 and 10 erl for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> service-class, respectively, distinguishing the traffic flow directions. The elastic service-classes behave as CBR services with the max BR (at the NDP only - not at the network operation phase (NOP)). We assume a 3% grade-of-service (end-to-end CBP) for each service-class achieved by applying the trunk reservation (TR) policy to the 1<sup>st</sup> and the 2<sup>nd</sup> service-class (with TR parameters of 1472 and 1152 Kbps, respectively) to benefit the 3<sup>rd</sup> service-class. The assigned bandwidth of 53.76 Mbps to each switching pair is calculated by the Erlang Multi Rate Loss Model (with TR) [7], and split to

38.40 and 15.36 Mbps for the shortest and the alternative path, respectively. The link capacities (see Fig. 2) are produced by summing up the assigned bandwidth of the paths passing through them. At the NOP, we use the TR parameters of 704 and 640 Kbps for the 1<sup>st</sup> and the 2<sup>nd</sup> service-class, respectively.

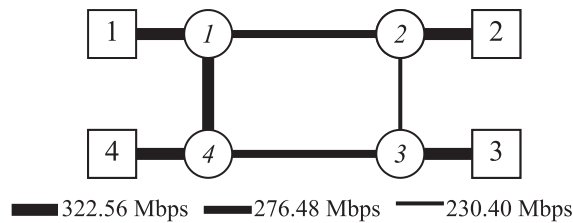


Fig. 2. Testbed network.

We have built a network simulator for simulating the call-level operations of connection-oriented networks [8]. Regarding the simulation of elastic service-classes, we model them as a concatenation of constant bit rate services, by exploiting the notion of “total offered traffic-load” (the product of: arrival rate x holding time x required bandwidth per call) [9]. For example, suppose an elastic service-class with  $b_{max} = 2$  Mbps (per call) and  $b_{min} = 1$  Mbps (per call). Let assume that such a call arrives at the network and initially asks to be serviced by its  $b_{max} = 2$  Mbps for 150 sec (holding time). At that time, the network cannot provide the bandwidth of  $b_{max}$ , but of 1.5 Mbps. Then the holding time of this call will be extended to 200 sec, according to the total offered traffic load principle (since  $2 * 150 = 1.5 * 200$ ).

CBP's presented herein are mean values of 6 simulation runs. The simulation is done according to the proposed CAC procedure. We consider 3 values of  $d$ , 1.8, 2.4 and 3 sec, for the a-MMF application (we borrow these values from the possible dial-tone delay in telephony [9]). We choose the scenario **c** which is the best as far as the resultant CBP is concerned since it gives more opportunities to a call to be conveyed than **b** or **a** (similarly **b** is better than **a**) [10]. Table I comparatively presents the mean CBP of the 3<sup>rd</sup> service-class for the three scenarios and justifies the choice of **c** (lower CBP). The max CBP of the network that occurs for the 3<sup>rd</sup> service-class is shown in Fig. 3 for  $d = 3$  sec, with or without the a-MMF (no FBA policy at all). The latter means that an arriving call immediately asks for call setup with a bandwidth between  $b_{min}$  and  $b_{max}$  (inclusive); blocking occurs when the  $avl$  is less than the  $b_{min}$ . The design-offered-traffic-load (corresponding to 0% of traffic fluctuation) fluctuates randomly among the switching pairs according to the uniform distribution by a max of  $x\%$  (where  $x = 20, 40, \dots, 100$  - horizontal axis of Fig. 3) [7]. This type of fluctuation represents the “complete random” fluctuation between the O-D pairs, where not only the traffic fluctuation is random (in some O-D pairs the traffic increases, while in some other O-D pairs the traffic decreases), but also the choice of O-D paths where the traffic fluctuation occurs. Fig. 3 shows significant improvement of the max CBP when the a-MMF is applied. In Table II we present the CBP of the 3<sup>rd</sup> service-class for the three

Comparison of the three different scenarios

Table 1

based on the CBP.

Traffic Fluctuation (%)	Mean CBP		
	Scenario a	Scenario b	Scenario C
0	0.052	0.051	0.049
20	0.091	0.089	0.086
40	0.164	0.160	0.155
60	0.227	0.225	0.216
80	0.281	0.275	0.263
100	0.322	0.318	0.302

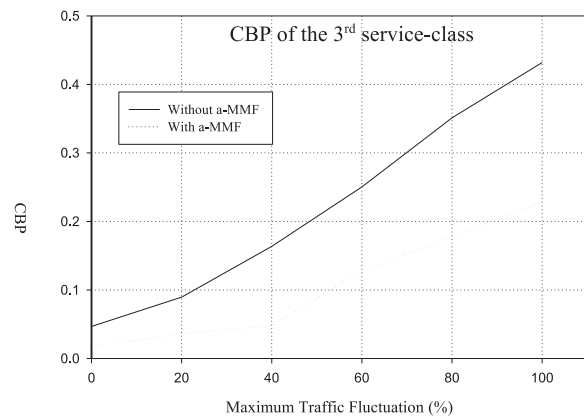


Fig. 3. Maximum CBP with and without a-MMF policy.

Maximum CBP with a-MMF policy versus  $d$ .

Table 2

Traffic Fluctuation	Delay $d=1.8$ sec CBP (%)	Delay $d=2.4$ sec CBP (%)	Delay $d=3$ sec CBP (%)
0%	$1.7 \pm 0.30$	$1.7 \pm 0.42$	$1.7 \pm 0.29$
20%	$3.7 \pm 0.48$	$3.7 \pm 0.33$	$3.7 \pm 0.48$
40%	$4.9 \pm 0.43$	$4.8 \pm 0.37$	$4.7 \pm 0.45$
60%	$13.0 \pm 1.32$	$12.9 \pm 0.93$	$12.8 \pm 1.06$
80%	$18.1 \pm 1.38$	$18.1 \pm 1.64$	$17.7 \pm 1.53$
100%	$23.2 \pm 1.60$	$22.6 \pm 0.48$	$22.6 \pm 1.57$

values of  $d$ . The last-column values of Table II coincide with CBP shown in Fig. 3. The confidence interval indicated in Table II is of 95%. We expected a decrease in CBP when  $d$  increases because more calls can take part in FBA each time the a-MMF is applied, utilizing better the  $avl$ , thus reducing the CBP. However, the results show that this happens only for very large traffic fluctuation. The 2<sup>nd</sup> service-class behaves similarly to the 3<sup>rd</sup> service-class, while the 1<sup>st</sup> service-class (CBR) behaves contrary to the elastic service-classes (Fig. 4). That is, when we get the best CBP for the elastic service-classes, we get the worst CBP for the CBR service-class.

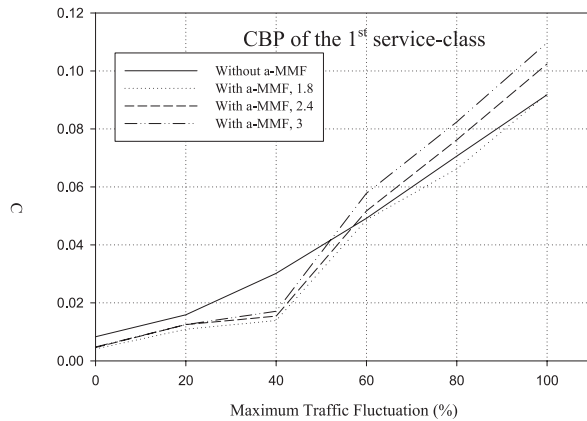


Fig. 4. Influence of the elastic traffic on the CBR traffic.

## 5. Conclusion

We propose a CAC procedure whereby we apply an FBA policy to the entire network. The critical part of the proposed CAC procedure is the definition of the delay. However, the high computational power of today's computer systems advocates a successful implementation of the proposed procedure. We delay calls at call setup for  $d$  sec, so as to apply the a-MMF policy. The available bandwidth of an O-D path is shared among calls of the same destination, whereas the bandwidth of in-service calls is not modified. Simulation results show significant lower CBP in the presence of the a-MMF policy. That is, the new CAC scheme not only achieves FBA but also improves substantially the call-level network performance. The influence of the delay  $d$  on CBP is small.

## References

- [1] HOU, Y. T., LI, B., PANWAR, S. S., TZENG, H.: *On network bandwidth allocation policies and feedback control algorithms for packet networks*, Computer Networks 34 (2000), pp. 481-501.
- [2] FAHMY, S., JAIN, R., KALYANARAMAN, S., GOYAL, R., VANDALORE, B.: *On Determining the Fair Bandwidth Share for ABR Connections in ATM Networks*, in Proc. ICC'98, June 1998, pp. 1485 - 1491.
- [3] MA, Q., STEENKISTE, P., ZHANG, H.: *Routing High-bandwidth Traffic in Max-Min Fair Share Networks*, in ACM Proc. SIGCOM'96, Stanford, CA, August 1996, pp. 206-217.
- [4] MOSCHOLIOS, I., LOGOTHETIS, M., KOKKINAKIS, G.: *A Parametric Linear Programming Model Describing Bandwidth Sharing Policies for ABR Traffic*, in Proc. 8th International Conference on Advances in Communication & Control, COMCON 8, Rethymna, Crete/Greece, June, 2001.
- [5] WAGNER, H. M.: *Principles of Operations Research*. Prentice-Hall: Englewood Cliffs, New Jersey 1969.
- [6] KESHAV, S.: *An Engineering Approach to Computer Networking*. Addison-Wesley: Massachusetts 1998.
- [7] LOGOTHETIS, M., SHIODA, S.: *Medium-term centralized virtual path bandwidth control based on traffic measurements*, IEEE Trans. Commun., vol. 43, pp. 2630-2640, October 1995.
- [8] LOGOTHETIS, M., LIOTOPOULOS, F.: *A Batch-type, Time-true ATM Network Simulator - Design for Parallel Processing*, International Journal of Communications Systems, Vol. 15, 8, pp. 713-739, October 2002.
- [9] AKIMARU, H., KAWASHIMA, K.: *Teletraffic*, Springer-Verlag: 2<sup>nd</sup> edition, Berlin, 1999.
- [10] LOGOTHETIS, M., MOSCHOLIOS, I., KOTZINOS, N.: *Network Simulation with Fair Bandwidth Allocation Policy and Centralized Path Bandwidth Control for Elastic Traffic*, in Proc. SoftCOM 2002, pp. 85-89.