Efficiency-Fairness Tradeoff in Telecommunications Networks

Moshe Zukerman, *Senior Member, IEEE,* Liansheng Tan, Hanwu Wang, and Iradj Ouveysi

***Abstract*— Introducing the concept of α-fairness, which allows for a bounded fairness compromise, so that no source is allocated less than a fraction α of its fair share, this letter studies tradeoffs between efficiency (utilization, throughput or revenue) and fairness in a general telecommunications network with relation to any fairness criterion. We formulate a linear program that finds the optimal bandwidth allocation by maximizing efficiency subject to α-fairness constraints. This leads to what we call an efficiency-fairness curve, which shows the benefit in efficiency as a function of the extent to which fairness is compromised.**

***Index Terms*— α-fairness, fairness criteria, efficiency-fairness tradeoff, bandwidth allocation.**

I. Introduction

Efficiency-fairness tradeoffs are prevalent in many aspects of life. Different societies and countries make their choices on these tradeoffs. This paper focuses on such tradeoffs in the context of telecommunications networks and provides a framework for evaluation and presentation of such tradeoffs. It applies to any network, topology and any fairness criterion. Possible applications include: Resilient Packet Ring (RPR) [4] [8] (the IEEE 802.17 standard for metropolitan area networks), local area networks (wireline and wireless) and wide area networks. Possible fairness criteria include: *RIAS fairness* [5] [6], *max-min fairness* [2] [3] [11], *proportional fairness* [9] [10], *general weighted* (GW) *fairness* [12] [13].

We introduce the concept of α-*fairness*. In particular, we define capacity assignment to be α-*fair* if no flow receives less than α times its fair allocation for $0 \leq \alpha \leq 1$. An interesting question is how to maximize *efficiency*, which can be defined as revenue, utilization, or throughput etc., under constraint of, say, 90% fairness. Another interesting question that we will be able to answer is how much efficiency can be improved by compromising on fairness to a certain extent.

Throughout the paper we use the notation $\langle x, y\rangle$ for the link that connects nodes x and y and $[x, y]$ for the data flow from node x to node y.

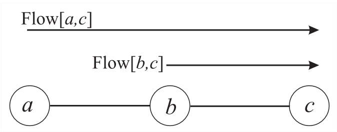

Fig. 1. Topology and flows of the three-node bus example.

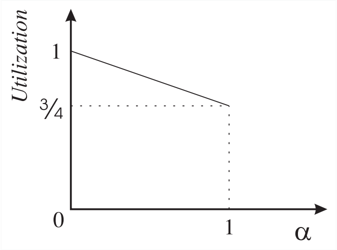

Fig. 2. The efficiency-fairness curve for the three-node bus example.

As an illustration, consider the three-node bus network presented in Fig. 1. All links are assumed identical in length and capacity and the capacity on each link is unity. Node a aims to transmit at unlimited rate to node c (designated as flow $[a, c]$), and node b also tries to transmit at unlimited rate to node c (designated as flow $[b, c]$). Assume that we choose max-min as our fairness criterion. Accordingly, each of the flows $[a, c]$ and $[b, c]$ will be assigned a rate of 1/2. Assuming all links are equal, this will mean that the utilization is 3/4 as link $\langle b, c\rangle$ is fully utilized and link $\langle a, b\rangle$ is only half utilized. However, if we relax the fairness constraint to α-fairness then flow $[b, c]$ will be assigned the rate of $\alpha/2$ and flow $[a, c]$ will be assigned the rate of $1 - \alpha/2$. This will give utilization of $1-\alpha/4$ plotted in Fig. 2. Full utilization is achieved if fairness is completely ignored ($\alpha = 0$) and flow $[a, c]$ is assigned the full bus capacity.

Clearly, the slope of the efficiency-fairness curve, for the previous example, can be made significantly steeper if efficiency is measured by revenue and flow $[a, c]$ pays much more than flow $[b, c]$ per bit/second transmitted per link. Notice that utilization can be viewed as a special case of revenue by setting the cost of the total capacity of each link to unity and the cost of each flow per link is equal to the proportion of the link capacity it uses. From now on, we will consider only revenue as a measure of efficiency.

Having introduced the concepts of α-fairness and the efficiency-fairness curve, in the remainder of this letter, we formulate a linear program that finds the optimal bandwidth allocation for a general network under the constraints of α-fairness and demonstrate its use and importance.

Manuscript received November 23, 2004. The associate editor coordinating the review of this letter and approving it for publication was Prof. Iakovos Venieris. This research has been partially supported by the National Natural Science Foundation of China under Grant No. 60473085 and partially by the Australian Research Council.

M. Zukerman is with the Australian Research Council Special Research Centre for Ultra-Broadband Information Networks (CUBIN), an affiliated program of National ICT Australia, Electrical and Electronic Engineering (EEE) Department, The University of Melbourne, Victoria 3010, Australia (e-mail: mzu@unimelb.edu.au).

L. Tan and H. Wang are with the Computer Science Department of Central China Normal University, Wuhan 430079, PR China (e-mail: l.tan@mail.ccnu.edu.cn).

I. Ouveysi is an honorary fellow of the EEE Department, The University of Melbourne, Victoria 3010, Australia (email: iradjouveysi@yahoo.com.uk).

Digital Object Identifier 10.1109/LCOMM.2005.07024.

II. The Optimization

We consider an N node network. The nodes are designated $1, 2, 3, \cdots, N$. All sources are assumed to be greedy. Let R_{ij} be the rate assigned to flow $[i, j]$. The aim is to set the R_{ij} values such that an objective function such as revenue or utilization is maximized subject to fairness and capacity constraints.

Let $P_{ij}\langle m, n\rangle$ be the price per bit/second transmitted on link $\langle m, n\rangle$ by flow $[i, j]$. Assume a unique route between any two end-points and let L_{ij} be the set of links in the route from i to j. Let F be the set of all flows. Assuming that the revenue obtained from flow $[i, j]$ grows linearly with its rate, the total revenue obtained by the network per second for transmission of flow $[i, j]$ is equal to

$$R_{ij} \times \sum_{\langle m,n\rangle \in L_{ij}} P_{ij}\langle m, n\rangle.$$

Therefore, the total revenue per second obtained by the network is given by

$$\sum_{[i,\, j]\, \in\, F} R_{ij} \sum_{\langle m,n\rangle \in L_{ij}} P_{ij}\langle m, n\rangle.$$

Defining the prices per link and per flow as above covers a wide range of pricing schemes. For example, if a pricing scheme is based on sources paying according to throughput, only the first link in a route may incur a positive cost. We acknowledge that pricing scheme may not be linear in which case other algorithms will be required to evaluate the efficiency-fairness curve.

Let $f(i, j)$ be the fair allocation for flow $[i, j]$ according to our chosen fairness criterion. Let us assume that we require that the allocation is α-fair. Then, R_{ij} will be bounded below by $\alpha f(i, j)$. This leads to the following linear programming formulation.

$$\text{Max} \quad E = \sum_{[i,\, j]\, \in\, F} R_{ij} \sum_{\langle m,n\rangle \in L_{ij}} P_{ij}\langle m, n\rangle$$

subject to

$$\begin{aligned} & R_{ij} \geq \alpha f(i, j) \\ & \sum_{[i,\, j]\, \in\, F\langle m,\, n\rangle} R_{ij} \leq C\langle m, n\rangle \qquad \text{for each link } \langle m, n\rangle \end{aligned}$$

where $C\langle m, n\rangle$ denotes the capacity of link $\langle m, n\rangle$ and $F\langle m, n\rangle$ is the set of all flows that use link $\langle m, n\rangle$. This optimization problem gives rise to the function $E(\alpha)$ which is our efficiency-fairness curve. This function can be obtained by using well known Linear Programming techniques [1]. All that is required is one run of the Simplex algorithm for, say, $\alpha = 1$ and one Simplex pivot per each segment of $E(\alpha)$. It is known from Chapter 6 of [1] that the function $E(\alpha)$ is piecewise-linear and concave. The end-points of each of its segments correspond to the values of α at which alternative optimal solutions exist for E.

III. Illustrating Examples

In this section, we will demonstrate the application of our linear programming formulation to the two examples illustrated in figures 3 and 4. RIAS fairness is assumed to be used in both examples. For the three-node ring of Fig. 3, we assume that there are three flows: $[1, 3]$, $[2, 1]$, and $[2, 3]$. These three flows are competing for the bandwidth on link $\langle 2, 3\rangle$.

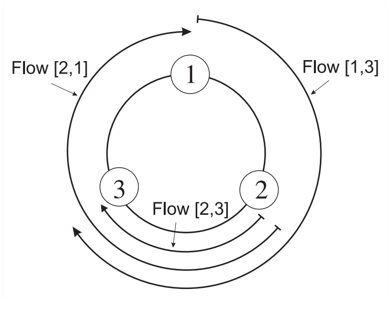

Fig. 3. Topology and flows for the three-node ring example.

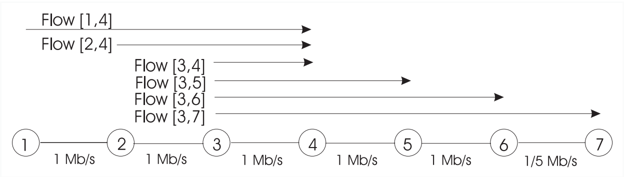

Fig. 4. Topology and flows for the seven-node bus example.

Assume that the capacities on links $\langle 1, 2\rangle$, $\langle 2, 3\rangle$ and $\langle 3, 1\rangle$ are 1 Mb/s, 2 Mb/s and 1.5 Mb/s, respectively. Let $P_{13}\langle 1, 2\rangle =$ 1 cents/Mb/s, $P_{13}\langle 2, 3\rangle = 2$ cents/Mb/s, $P_{23}\langle 2, 3\rangle = 3$ cents/Mb/s, $P_{21}\langle 2, 3\rangle = 3$ cents/Mb/s, $P_{21}\langle 3, 1\rangle = 2$ cents/Mb/s. According to the RIAS fairness: $f(1, 3) = 1$ Mb/s, $f(2, 3) = 0.5$ Mb/s and $f(2, 1) = 0.5$ Mb/s. In this example, the efficiency/revenue function can be written as

$$E = 3R_{13} + 3R_{23} + 5R_{21}.$$

The efficiency-fairness curve is thus obtained by the following linear programming formulation.

$$\text{Max} \quad E = 3x_1 + 3x_2 + 5x_3$$

subject to

$$\begin{aligned} & x_1 \geq \alpha f(1, 3) = \alpha \\ & x_2 \geq \alpha f(2, 3) = 0.5\alpha \\ & x_3 \geq \alpha f(2, 1) = 0.5\alpha \\ & x_1 + x_2 + x_3 \leq 2, \\ & x_1 \leq 1, \\ & x_3 \leq 1.5 \\ & x_i \geq 0 \text{ for } i = 1, 2, 3 \end{aligned}$$

where $x_1 = R_{13}$, $x_2 = R_{23}$, and $x_3 = R_{21}$.

For the model of Fig. 4, the total cost (end-to-end) of each of the flows excluding [3,7] is assumed to be 1 cent/Mb/s. The total cost (end-to-end) of flow [3,7] is 48 cents/Mb/s. Therefore, the efficiency/revenue function can be written as

$$E = R_{14} + R_{24} + R_{34} + R_{35} + R_{36} + 48R_{37}.$$

The efficiency-fairness curve is thus obtained by the following.

$$\text{Max} \quad E = x_1 + x_2 + x_3 + x_4 + x_5 + 48x_6$$

subject to

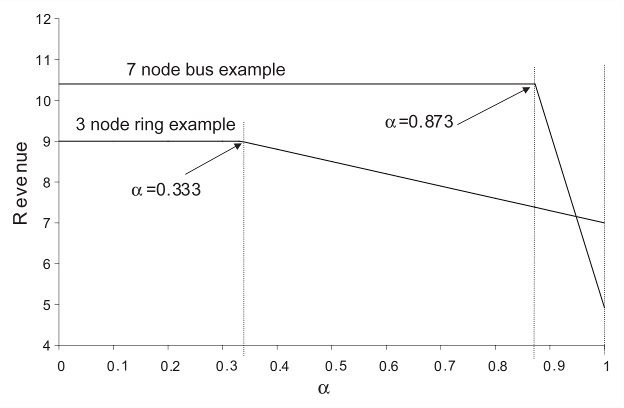

Fig. 5. The efficiency/revenue-fairness curves for the three-node ring and the seven-node bus examples.

$$\begin{aligned}
&x_1 \geq \alpha f(1,4) = \alpha/3\\
&x_2 \geq \alpha f(2,4) = \alpha/3\\
&x_3 \geq \alpha f(3,4) = \alpha/12\\
&x_4 \geq \alpha f(3,5) = \alpha/12\\
&x_5 \geq \alpha f(3,6) = \alpha/12\\
&x_6 \geq \alpha f(3,7) = \alpha/12\\
&x_1 \leq 1\\
&x_1 + x_2 \leq 1\\
&x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \leq 1\\
&x_4 + x_5 + x_6 \leq 1\\
&x_5 + x_6 \leq 1\\
&x_6 \leq \tfrac{1}{5}\\
&x_i \geq 0 \text{ for } i = 1,2,3,4,5,6
\end{aligned}$$

where $x_1 = R_{13}$, $x_2 = R_{24}$, $x_3 = R_{34}$, $x_4 = R_{35}$, $x_5 = R_{36}$, and $x_6 = R_{37}$. In Fig. 5 we present the efficiency/revenue-fairness ($E(\alpha)$) functions for the two examples.

These two examples produce efficiency-fairness curves that are different in the slope of the right-hand segment. The seven-node bus example clearly produces a much steeper slope for that right-hand segment, meaning that giving up a small percentage of fairness can lead to a significant increase in revenue. Moreover, the seven-node bus example demonstrates that there are cases whereby after trading significant increase in revenue for a small percentage of fairness, no further revenue increase is possible even if further compromise on fairness is made. Such considerations can be made by designers using the efficiency-fairness curve.

IV. Discussion

The concept of efficiency-fairness tradeoff in telecommunications networking and standardization is not new. Some of us may remember the heated debate over the fairness of the Distributed Queue Dual Bus (DQDB) Metropolitan Area Network standard (IEEE 802.6) [7], [14] which includes the so-called Bandwidth Balancing option that allows operators to control the utilization-fairness tradeoff.

What is new in this paper is the α-fairness concept and with it the efficiency-fairness curve that allow designers and standard developers to choose the level of fairness and to consider other important economical factors. We propose that standard committees will accept a certain level of flexibility when they set fairness requirements. After all, fairness in networks applies only during congestion periods, and in many cases users may be happy to compromise on some fairness for a lower service cost.

V. Conclusions

This letter studies tradeoffs between efficiency (utilization or revenue) and fairness in a general telecommunications network with relation to any fairness criterion. We formulate a linear programming model that finds the optimal bandwidth allocation for a general network under the constraints of α-fairness. This leads to what we call efficiency-fairness curves, which shows the benefit in efficiency as a function of the extent to which fairness is compromised. We use two examples based on a three-node ring and a seven-node bus network to demonstrate the design implication of our efficiency-fairness curve. The new concepts presented here may be applicable to other social and economic fields.

References

[1] M. S. Bazaraa and J. J. Jarvis, *Linear Programming and Network Flows*. New York: John Wiley & Sons, 1976.

[2] D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall, 1992.

[3] A. Charny, D. D. Clark, and R. Jain, “Congestion control with explicit rate indication,” in *Proc. IEEE International Conference on Communications*, vol. 3, June 1995, pp. 1954-1963.

[4] F. Davik, M. Yilmaz, S. Gjessing, and N. Uzun, “IEEE 802.17 resilient packet ring tutorial,” *IEEE Commun. Mag.*, vol. 42, pp. 112-118, Mar. 2004.

[5] V. Gambiroza, Y. Liu, P. Yuan, and E. Knightly, “High performance fair bandwidth allocation for resilient packet rings,” in *Proc 15th ITC Specialist Seminar on Traffic Engineering and Traffic Management*, July 2002.

[6] V. Gambiroza, P. Yuan, L. Balzano, Y. Liu, S. Sheafor, and E. Knightly, “Design, analysis, and implementation of DVSR: a fair high performance protocol for packet rings,” *IEEE/ACM Trans. Networking*, vol. 12, pp. 85-102, Feb. 2004.

[7] E. L. Hahne, A. K. Choudhury, and N. F. Maxemchuk, ”DQDB networks with and without bandwidth balancing,” *IEEE Trans. Commun.*, vol. 40, pp. 1192-1204, July 1992.

[8] C. N. Hawkins, J. Green, M. Sharma, and K. Vasani, “Resilient packet rings for metro networks,” Aug. 2001 (http://www.rpralliance.org/).

[9] F. Kelly, “Charging and rate control for elastic traffic,” *European Trans. Telecommun.*, vol. 8, pp. 33-37, Jan. 1997.

[10] L. Massoulie and J. Roberts, “Bandwidth sharing: objectives and algorithms,” in *Proc. 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, Mar. 1999, pp. 1395-1403.

[11] L. Tan, A. C. Pugh, and M. Yin, “Rate-based congestion control in ATM switching networks using a recursive digital filter”, *Control Engineering Practice*, vol. 11, pp. 1171-1181, Oct. 2003.

[12] B. Vandalore, S. Fahmy, R. Jain, R. Goyal, and M. Goyal, “A definition of general weighted fairness and its support in explicit rate switch algorithms,” in *Proc. Sixth International Conference on Network Protocols (ICNP)*, Oct. 1998, pp. 22-30.

[13] B. Vandalore, S. Fahmy, R. Jain, R. Goyal, and M. Goyal, “General weighted fairness and its support in explicit rate switch algorithms,” *Computer Commun.*, vol. 23, pp. 149-161, Jan. 2000.

[14] M. Zukerman and P. Potter, ”The DQDB protocol and its performance under overload traffic conditions,” *Computer Network and ISDN Systems*, vol. 20, pp. 261-270, Dec. 1990.