

Report of the Institute of Control and Computation Engineering Warsaw University of Technology

A Multi-Criteria Approach to Fair and Efficient Bandwidth Allocation

Włodzimierz Ogryczak, Adam Wierzbicki and Marcin Milewski

February, 2005

Report nr: 05-02

Copyright ©2005 by the Institute of Control & Computation Engineering

Permission to use, copy and distribute this document for any purpose and without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies, and that the name of ICCE not be used in advertising or publicity pertaining to this document without specific, written prior permission. ICCE makes no representations about the suitability of this document for any purpose. It is provided “as is” without express or implied warranty.



Warsaw University of Technology
Institute of Control & Computation Engineering
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland
phone (48-22)-660-7397, fax (48-22)-825-3719



A Multi-Criteria Approach to Fair and Efficient Bandwidth Allocation

Włodzimierz Ogryczak*
ogryczak@ia.pw.edu.pl

Adam Wierzbicki†
adamw@icm.edu.pl,

Marcin Milewski‡
mmilewsk@elka.pw.edu.pl,

February, 2005

Abstract

In systems which serve many users there is a need to respect some fairness rules while looking for the overall efficiency. This applies among others to networking where a central issue is how to allocate bandwidth to flows efficiently and fairly. The so-called Max-Min Fairness is widely used to meet these goals. Although, allocating the bandwidth to optimize the worst performances may cause a large worsening of the overall throughput of the network. In this paper we show how the concepts of multiple criteria equitable optimization can be effectively used to generate various fair and efficient allocation schemes. We introduce a multiple criteria model equivalent to equitable optimization and we develop a corresponding reference point procedure to generate fair and efficient bandwidth allocations. Our analysis is focused on the nominal network design for elastic traffic which becomes recently increasingly significant. The procedure is tested on a sample network dimensioning problem for elastic traffic and its abilities to model various preferences are demonstrated.

Key words. Multiple Criteria Optimization, Efficiency, Fairness, Equity, Reference Point Method, Resource Allocation, Telecommunications, Network Design, Elastic Traffic Telecommunications,

1 Introduction

Resource allocation decisions are usually concerned with the allocation of limited resources so as to achieve the best system performances. However, in networking there is a need to respect some fairness rules while looking for the overall efficiency. A fair way of distribution of the bandwidth (or other network resources) among competing demands becomes a key issue in computer networks [3]

*Partial support provided by The State Committee for Scientific Research under grant 3T11C 005 27 “Models and Algorithms for Efficient and Fair Resource Allocation in Complex Systems”.

†Polish-Japanese Institute of Information Technology; partial support provided by The State Committee for Scientific Research under grant 3T11C 005 27 “Models and Algorithms for Efficient and Fair Resource Allocation in Complex Systems”.

‡Institute of Telecommunications, Warsaw University of Technology; partial support provided The State Committee for Scientific Research under grant 3T11D 001 27 “Design Methods for NGI Core Networks”.

and telecommunications network design, in general [16, 20]. This paper deals with problems of bandwidth allocation within telecommunication networks. Therefore, we focus on the approaches that, while allocating resources, attempt to provide a fair (equal) treatment of all the activities (demands) [9, 13].

Expanding demand on the Internet services has led to an increased role of the traffic carried by the IP protocol in telecommunication networks. The TCP protocol is the most frequently used transport protocol in best-effort IP networks. The data traffic carried by the TCP protocol adapts its throughput to the amount of available bandwidth. Such traffic, called *elastic traffic*, is capable to use the entire available bandwidth, but it will also be able to reduce its throughput in the presence of contending traffic. Nowadays, the network management often faces the problem of designing networks that carry elastic traffic. These network design problems are, essentially, the network dimensioning problems as they can be reduced to a decision about link capacities. Flow sizes are outcomes of the design problem, since the flows adapt to given network resources on a chosen path.

Network management must stay within a budget constraint on link bandwidth to expand network capacities. They want to achieve a high throughput of the IP network, to increase the multiplexing gains (due to the use of packet switching by the IP protocol). This traffic is offered only a best-effort service, and therefore network management is not concerned with offering guaranteed levels of bandwidth to the traffic. A straightforward network dimensioning with elastic traffic could be thought of as a search for such network flows that will maximize the aggregate network throughput while staying within a budget constraint for the costs of link bandwidth. However, maximizing aggregate throughput can result in extremely unfair solutions allowing even for starvation of flows for certain services. On the other extreme, while looking at the problem from the perspective of a network user, the network flows between different nodes should be treated as fairly as possible [2]. The so-called Max-Min Fairness (MMF) [1, 4] is widely considered as such ideal fairness criteria. Indeed, the lexicographic max-min optimization used in the MMF approach generalizes equal sharing at a single link bandwidth to any network while maintaining the Pareto optimality. Certainly, allocating the bandwidth to optimize the worst performances may cause a large worsening of the overall throughput of the network. Therefore, network management must consider two goals: increasing throughput and providing fairness. These two goals are clearly conflicting, if the budget constraint has to be satisfied.

The search for compromise solutions that do not starve network flows, and give satisfying levels of throughput has led to the development of methods depending on maximization of the sum of the flows evaluated with some (concave) utility function. In particular, the so-called *Proportional Fairness* approach [5] maximizes the sum of logarithms of the flows. The approach has been further extended to a parametric class of concave utility functions [11]. However, every such approach requires to build (or to guess) a utility function prior to the analysis and later it gives only one possible compromise solution. It is very difficult to identify and formalize the preferences at the beginning of the decision process. Therefore, a decision support process is usually needed which attempts to gain additional preference information interactively, allowing simultaneously the decision-maker (DM) to learn the problem during the process with possibly evolving preferences. This can be effectively achieved with the so-called quasi-satisficing approach to multiple criteria decision problems [22]. The best formalization of the quasi-satisficing approach to multiple criteria optimization was proposed and developed mainly by Wierzbicki [21] as the reference point method. The reference point method (RPM) is an interactive technique where the DM specifies preferences in terms of aspiration levels (reference point), i.e., by introducing desired (acceptable) levels for several criteria. The purpose of this work is to show that there exists a multiple criteria model that allows to represent consistently the overall efficiency and fairness goals and thereby to apply effectively the reference point methodology to the bandwidth allocation problems.

The paper is organized as follows. In the next section we formalize the network dimensioning prob-

lem, we consider. In Section 3, basic fair solution concepts for resource allocation are related to the multiple criteria equitable optimization theory. In Section 4, the reference point methodology is applied to the multiple criteria allowing us to model various fair and efficient allocation schemes with simple control parameters. Finally, in Section 5, we present some results of our initial computational experience with this new approach.

2 The bandwidth allocation problem

The problem of network dimensioning with elastic traffic can be formulated basically as a Linear Programming (LP) based resource allocation model as follows (Pióro and Medhi, 2004). Given a network topology $G = \langle V, E \rangle$, consider a set of pairs of nodes as the set $I = \{1, 2, \dots, m\}$ of services representing the elastic flow from source v_i^s to destination v_i^d . For each service, we have given the set P_i of possible routing paths in the network from the source to the destination. This information can be summarized in the form of binary matrices $\Delta_e = (\delta_{eip})_{i \in I; p \in P_i}$ assigned to each link $e \in E$, where $\delta_{eip} = 1$ if link e belongs to the routing path $p \in P_i$ (connecting v_i^s with v_i^d) and $\delta_{eip} = 0$ otherwise.

For each service $i \in I$, the elastic flow from source v_i^s to destination v_i^d is a variable representing the model outcome and it will be denoted by x_i . This flow may be realized along various paths $p \in P_i$. The flow may be either split among several paths or a single path must be finally selected to serve the entire flow. Actually, the latter case of nonbifurcated flows is more commonly required. Both bifurcated or nonbifurcated flows may be modeled as $x_i = \sum_{p \in P_i} x_{ip}$ where x_{ip} (for $p \in P_i$) are nonnegative variables representing the elastic flow from source v_i^s to destination v_i^d along the routing p . Although, the single-path model requires additional multiple choice constraints to enforce nonbifurcated flows. This can be implemented with additional binary (flow assignment) variables u_{ip} equal 1 if path $p \in P_i$ is assigned to serve flow x_i and 0 otherwise. Assuming existence of some constant M upper bounding the largest possible total flows x_i , the assignment variables u_{ip} can be easily used to limit the number of positive flows x_{ip} with the following constraints:

$$0 \leq x_{ip} \leq M u_{ip}, \quad u_{ip} \in \{0, 1\} \quad \forall i \in I; p \in P_i \quad (1)$$

$$\sum_{p \in P_i} u_{ip} = 1 \quad \forall i \in I \quad (2)$$

The network dimensioning problem depends on allocating the bandwidth to several links in order to maximize flows of all the services (demands). Typically, the network is already operated and some bandwidth is already allocated (installed) and decisions are rather related to the network expansion. Therefore, we assume that each link $e \in E$ has already capacity a_e while decision variables ξ_e represent the bandwidth newly allocated to link $e \in E$ thus expanding the link capacity to $a_e + \xi_e$. Certainly, all the decision variables must be nonnegative: $\xi_e \geq 0$ for all $e \in E$ and there are usually some bounds (upper limits) on possible expansion of the links capacities: $\xi_e \leq \bar{a}_e$ for all $e \in E$. Finally, the following constraints must be fulfilled:

$$\sum_{i \in I} \sum_{p \in P_i} \delta_{eip} x_{ip} \leq a_e + \xi_e \quad \forall e \in E \quad (3)$$

$$0 \leq \xi_e \leq \bar{a}_e \quad \forall e \in E \quad (4)$$

$$\sum_{p \in P_i} x_{ip} = x_i \quad \forall i \in I \quad (5)$$

where (5) define the total service flows, while (3) establish the relation between service flows and links bandwidth. The quantity $y_e = \sum_{i \in I} \sum_{p \in P_i} \delta_{eip} x_{ip}$ is the load of link e and it cannot exceed the available link capacity.

Further, for each link $e \in E$, the cost of allocated bandwidth is defined. In the basic model of network dimensioning it is assumed that any real amount of bandwidth may be installed and marginal costs c_e of link bandwidth is given. Hence, the corresponding link dimensioning function expressing amount of capacity (bandwidth) necessary to meet a required link load [16] is then a linear function. While allocating the bandwidth to several links in the network dimensioning process the decisions must keep the cost within available budget B for all link bandwidths. Hence the following constraint must be satisfied:

$$\sum_{e \in E} c_e \xi_e \leq B. \quad (6)$$

The model constraints (3)–(6) together with respective nonnegativity requirements define a linear programming (LP) feasible set. They turn into Mixed Integer LP (MILP), however, if nonbifurcated flows are enforced with discrete constraints (1)–(2).

Link modularity (bandwidth granulation) is a common feature in communications networks [16]. Therefore, in more realistic models, for each link $e \in E$, the minimum unit of bandwidth b_e is assumed to be available for allocation (installation) and c_e represents the corresponding unit cost. The corresponding link dimensioning function is then a step wise function. In the case of modular links (discrete bandwidth units b_e), the installed capacity ξ_e must satisfy additional equation:

$$\xi_e = b_e z_e \quad \forall e \in E \quad (7)$$

where z_e is an integer decision variable representing the number of bandwidth units installed at link e . The model constraints (3)–(6) extended with (7) turns then into MILP feasible set even if bifurcated flows are allowed.

In the simplified problem with linear link dimensioning function (no modularities) and dimensioning of a completely new network ($a_e = 0$ for all links), the cost of the entire path p for service i can be directly expressed by the formula:

$$\kappa_{ip} = \sum_{e \in E} c_e \delta_{eip} \quad \text{for } i = 1, \dots, m, p \in P_i. \quad (8)$$

The cheapest path for each service can be then easily identified and preselected. Having preselected routing path for each demand ($|P_i| = 1$) one may consider variable x_i directly as flow along the corresponding path ($x_i = x_{i1}$). Constraints (6) and (3) may be then treated as equations and together with formula (8) they allow one to eliminate variables ξ_e , thus formulating the problem as a simplified resource allocation model with only one constraint:

$$\sum_{i=1}^m \kappa_i x_i = B \quad \text{where } \kappa_i = \kappa_{i1} \quad \forall i \in I, \quad (9)$$

and variables x_i representing directly the decisions. Note that one cannot define directly any cost κ_{ip} of the path $p \in P_i$ (similar to (8)) when some capacity is already available ($a_e > 0$ for some $e \in E$). In other words in the problem, we consider, the cost of available link capacity is actually nonlinear (piecewise linear) and this results in the lack of direct formula for the path cost since it depends on possible sharing with other paths of the preinstalled bandwidth (free capacity a_e). Such a simplification is, certainly, also impossible for a modular case, due to additional discrete constraints (7).

The network dimensioning model can be considered with various objective functions, depending on the chosen goal. One may consider two extreme approaches. The first extreme is the maximization of the total throughput (the sum of flows) $\sum_{i \in I} x_i$. On the other extreme, the network flows between different nodes should be treated as fairly as possible which leads to the maximization of the smallest flow or rather to the lexicographically expanded max-min optimization (the so-called max-min ordering) allowing also to maximize the second smallest flows provided that the smallest remain optimal, the third smallest, etc. This approach is widely recognized in networking as the so-called Max-Min Fairness (MMF) [1, 4] and it is consistent with the Rawlsian theory of justice [17].

Note that in the simplified dimensioning model (with preselected paths and continuous bandwidth), due to possible alternative formulation (9), the throughput maximization approach apparently would choose one variable x_{i^0} which has the smallest marginal cost $\kappa_{i^0} = \min_{i \in I} \kappa_i$ and make that flow maximal within the budget limit ($x_{i^0} = B/\kappa_{i^0}$), while eliminating all other flows (lowering them to zero). On the other hand, the MMF concept applied to the simplified dimensioning model (resulting in (9)) would lead us to a solution with equal values for all the flows: $x_i = B/\sum_{i \in I} \kappa_i$ for $i \in I$. Such allocating the resources to optimize the worst performances may cause a large worsening of the overall (mean) performances as the MMF throughput ($mB/\sum_{i \in I} \kappa_i$) might be considerably smaller than the maximal throughput ($B/\min_{i \in I} \kappa_i$). In more realistic dimensioning models assuming bandwidth modularity or other nonlinearities in link dimensioning function (like the existence of a free capacity a_e of preinstalled bandwidth) and nonbifurcation requirements a direct formula for a path cost is not available and the corresponding solutions are not so clear. Nevertheless, the main weaknesses of the above solutions remain valid. The throughput maximization can always result in extremely unfair solutions allowing even for starvation of certain flows while the MMF solution may cause a large worsening of the throughput of the network. In an example built on the backbone network of a Polish ISP, it turned out that the throughput in a perfectly fair solution could be less than 50% of the maximal throughput [14].

Network management may be interested in seeking a compromise between the two extreme approaches discussed above. One of possible solutions depends on maximization of the sum of the flows evaluated with some (concave) utility function $\sum_{i \in I} U_i(x_i)$. A parametric class of utility functions [11]:

$$U_i(x_i, \alpha) = \begin{cases} x_i^{1-\alpha}/(1-\alpha) & \text{if } \alpha \neq 1 \\ \log(x_i) & \text{if } \alpha = 1 \end{cases}$$

may be used for this purpose generating various solution concepts for $\alpha \geq 0$. In particular, for $\alpha = 0$ one gets the throughput maximization which is the only linear criterion within the entire class. For $\alpha = 1$, it represents the Proportional Fairness approach [5] that maximizes the sum of logarithms of the flows while it converges to the MMF with α tending to the infinity. However, every such approach requires to build (or to guess) a utility function prior to the analysis and later it gives only one possible compromise solution. It is very difficult to identify and formalize the preferences at the beginning of the decision process. Moreover, apart from the trivial case of throughput maximization all the utility functions that really take into account any fairness preferences are nonlinear. Nonlinear objective functions applied to the MILP models we consider results in computationally hard optimization problems. In the following, we shall describe an approach that allows to search for such compromise solutions with multiple linear criteria rather than the use nonlinear objective functions.

3 Fair allocations and equitable efficiency

The bandwidth allocation problem, we consider, may be viewed as a special case of general resource allocation problem where a set I of m services is considered and for each service $i \in I$, its measure of realization x_i is a function $x_i = f_i(\xi)$ of the allocation pattern $\xi \in A$. This function, called the individual objective function, represents the outcome (effect) of the allocation pattern for service i . In applications, we consider, the measure expresses the service flow and a larger value of the outcome means a better effect (higher service quality or client satisfaction). This leads us to a vector maximization problem:

$$\max \{(x_1, x_2, \dots, x_m) : \mathbf{x} \in Q\} \quad (10)$$

where $Q = \{(x_1, \dots, x_m) : x_i = f_i(\xi) \text{ for } i \in I, \xi \in A\}$ denotes the attainable set for outcome vectors \mathbf{x} . For the network dimensioning problems, we consider, the set Q is an MILP feasible set defined by basic constraints (1)–(6) with additional discrete constraints (7) in the case of modular bandwidth.

Model (10) only specifies that we are interested in maximization of all outcomes x_i for $i \in I$. In order to make it operational, one needs to assume some solution concept specifying what it means to maximize multiple outcomes. The solution concepts are defined by properties of the corresponding preference model within the outcome space. The commonly used concept of the Pareto-optimal solutions, as feasible solutions for which one cannot improve any criterion without worsening another, depends on the rational dominance \succ_r which may be expressed in terms of the vector inequality: $\mathbf{x}' \succ_r \mathbf{x}''$ iff $x'_i \geq x''_i$ for all $i \in I$.

The concept of fairness has been studied in various areas beginning from political economics problems of fair allocation of consumption bundles to abstract mathematical formulation [18]. In order to ensure fairness in a system, all system entities have to be equally well provided with the system's services. This leads to concepts of fairness expressed by the equitable rational preferences [12, 6]. First of all, the fairness requires impartiality of evaluation, thus focusing on the distribution of outcome values while ignoring their ordering. That means, in the multiple criteria problem (10) we are interested in a set of outcome values without taking into account which outcome is taking a specific value. Hence, we assume that the preference model is impartial (anonymous, symmetric). In terms of the preference relation it may be written as the following axiom

$$(x_{\tau(1)}, x_{\tau(2)}, \dots, x_{\tau(m)}) \cong (x_1, x_2, \dots, x_m) \quad (11)$$

for any permutation τ of I . Further, fairness requires equitability of outcomes which causes that the preference model should satisfy the (Pigou–Dalton) principle of transfers. The principle of transfers states that a transfer of any small amount from an outcome to any other relatively worse-off outcome results in a more preferred outcome vector. As a property of the preference relation, the principle of transfers takes the form of the following axiom

$$x_{i'} > x_{i''} \quad \Rightarrow \quad \mathbf{x} - \varepsilon \mathbf{e}_{i'} + \varepsilon \mathbf{e}_{i''} \succ \mathbf{x} \quad \text{for } 0 < \varepsilon < x_{i'} - x_{i''} \quad (12)$$

The rational preference relations satisfying additionally axioms (11) and (12) are called hereafter *fair (equitable) rational preference relations*. We say that outcome vector \mathbf{x}' *fairly dominates* \mathbf{x}'' ($\mathbf{x}' \succ_e \mathbf{x}''$), iff $\mathbf{x}' \succ \mathbf{x}''$ for all fair rational preference relations \succ . An allocation pattern $\xi \in A$ is called *fairly (equitably) efficient* if $\mathbf{x} = \mathbf{f}(\xi)$ is fairly nondominated. Note that each fairly efficient solution is also Pareto-efficient, but not vice versa.

The theory of majorization [10] includes the results which allow us to express the relation of fair (equitable) dominance as a vector inequality on the cumulative ordered outcomes [6]. This can be

mathematically formalized as follows. First, introduce the ordering map $\Theta : R^m \rightarrow R^m$ such that $\Theta(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \dots, \theta_m(\mathbf{x}))$, where $\theta_1(\mathbf{x}) \leq \theta_2(\mathbf{x}) \leq \dots \leq \theta_m(\mathbf{x})$ and there exists a permutation τ of set I such that $\theta_i(\mathbf{x}) = x_{\tau(i)}$ for $i = 1, \dots, m$. Next, apply to ordered outcomes $\Theta(\mathbf{x})$, a linear cumulative map thus resulting in the *cumulative ordering map* $\bar{\Theta}(\mathbf{x}) = (\bar{\theta}_1(\mathbf{x}), \bar{\theta}_2(\mathbf{x}), \dots, \bar{\theta}_m(\mathbf{x}))$ defined as

$$\bar{\theta}_i(\mathbf{x}) = \sum_{j=1}^i \theta_j(\mathbf{x}) \quad \text{for } i = 1, \dots, m \quad (13)$$

The coefficients of vector $\bar{\Theta}(\mathbf{x})$ express, respectively: the smallest outcome, the total of the two smallest outcomes, the total of the three smallest outcomes, etc. The theory of majorization allow us to derive the following theorem [6].

THEOREM 3.1 *Outcome vector \mathbf{x}' fairly dominates \mathbf{x}'' , if and only if $\bar{\theta}_i(\mathbf{x}') \geq \bar{\theta}_i(\mathbf{x}'')$ for all $i \in I$ where at least one strict inequality holds.*

Note that Theorem 3.1 permits one to express fair solutions of problem (10) as Pareto-efficient solutions to the multiple criteria problem with objectives $\bar{\Theta}(\mathbf{x})$

$$\max \{(\eta_1, \eta_2, \dots, \eta_m) : \eta_i = \bar{\theta}_k(\mathbf{x}) \quad \text{for } k = 1, \dots, m, \mathbf{x} \in Q\} \quad (14)$$

For better understanding of the multiple criteria problem (14), one may consider normalized objective functions $\mu_k(\mathbf{x}) = \bar{\theta}_k(\mathbf{x})/k$ thus representing for each k the mean of the k smallest outcomes, called the worst conditional mean [13]. Note that the last (m -th) objective in problem (14) represents the sum of outcomes thus corresponding to throughput maximization. Simple maximin optimization corresponds to maximization of the first objective in (14). The complete MMF solution concept represents the lexicographic approach to multiple criteria in (14):

$$\text{lexmax} \{(\eta_1, \eta_2, \dots, \eta_m) : \eta_i = \bar{\theta}_k(\mathbf{x}) \quad \text{for } k = 1, \dots, m, \mathbf{x} \in Q\}$$

While the MMF is only a specific (extreme) solution concept, the entire multiple criteria problem (14) may serve as a source of various fairly efficient allocation schemes. Moreover, although the definitions of quantities $\bar{\theta}_k(\mathbf{x})$ are very complicated, they can be modeled with simple auxiliary constraints. Let us notice that for any given vector \mathbf{x} , the quantity $\bar{\theta}_k(\mathbf{x})$ is defined by the following LP problem:

$$\begin{aligned} \bar{\theta}_k(\mathbf{x}) &= \min \sum_{i=1}^m x_i u_{ki} \\ \text{s.t.} \quad &\sum_{i=1}^m u_{ki} = k, \quad 0 \leq u_{ki} \leq 1 \quad \text{for } i = 1, \dots, m. \end{aligned} \quad (15)$$

Exactly, the above problem is an LP for a given outcome vector \mathbf{x} while it begins nonlinear for a variable \mathbf{x} . This difficulty can be overcome by taking advantages of the LP dual to (15):

$$\begin{aligned} \bar{\theta}_k(\mathbf{x}) &= \max \quad kt - \sum_{i=1}^m d_i \\ \text{s.t.} \quad &t - x_i \leq d_i, \quad d_i \geq 0 \quad \text{for } i = 1, \dots, m \end{aligned} \quad (16)$$

where t is an unrestricted variable while nonnegative variables d_i represent, for several outcome values x_i , their downside deviations from the value of t [15].

Formula (16) allows us to formulate problem (14) as follows:

$$\max (\eta_1, \eta_2, \dots, \eta_m) \quad (17)$$

$$\text{subject to } \mathbf{x} \in Q$$

$$\eta_k = kt_k - \sum_{i=1}^m d_{ik} \quad \text{for } k = 1, \dots, m \quad (18)$$

$$t_k - d_{ik} \leq x_i, \quad d_{ik} \geq 0 \quad \text{for } i, k = 1, \dots, m \quad (19)$$

Note that problem (17)–(19) adds only linear constraints to the original attainable set Q . Hence, for the basic network dimensioning problems with the set Q defined by constraints (1)–(6), the resulting formulation (17)–(19) remains in the class of (multiple criteria) MILP. The same applies to the modular dimensioning model with additional constraints (7). For the simplified LP model (3)–(6) with flows bifurcation allowed and continuous bandwidth the multiple criteria formulation (17)–(19) remains in the class of (multiple criteria) LP.

Although defined with simple linear constraints, the expanded model (17)–(19) introduces m^2 additional variables and inequalities. This may cause a serious computational burden for real-life network dimensioning problems. Note that the number of services (traffic demands) corresponds to the number of ordered pairs of network nodes which is already square of the number of nodes $|V|$. Thus, finally the expanded multiple criteria model introduces $|V|^4$ variables and constraints which means polynomial but fast growth and can be not acceptable for larger networks. For instance, rather small backbone network of Polish ISP [14], we analyze in Section 5, consists of 12 nodes which leads to 132 elastic flows ($m = 132$) resulting in 17 424 constraints (19) and the same number of deviational variables d_{ik} . In order to reduce the problem size one may attempt the restrict the number of criteria in the problem (14).

Let us consider a sequence of indices $K = \{k_1, k_2, \dots, k_q\}$, where $1 = k_1 < k_2 < \dots < k_{q-1} < k_q = m$, and the corresponding restricted form of the multiple criteria model (14):

$$\max \{(\eta_{k_1}, \eta_{k_2}, \dots, \eta_{k_q}) : \eta_k = \bar{\theta}_k(\mathbf{x}) \quad \text{for } k \in K, \quad \mathbf{x} \in Q\} \quad (20)$$

with only $q < m$ criteria. Following Theorem 3.1, multiple criteria model (14) allows us to generate any fairly efficient solution of problem (10). Reducing the number of criteria we restrict these opportunities. Nevertheless, one may still generate reasonable compromise solutions. First of all the following assertion is valid.

THEOREM 3.2 *If \mathbf{x}^o is an efficient solution of the restricted problem (20), then it is an efficient (Pareto-optimal) solution of the multiple criteria problem (10) and it can be fairly dominated only by another efficient solution \mathbf{x}' of (20) with exactly the same values of criteria: $\bar{\theta}_k(\mathbf{x}') = \bar{\theta}_k(\mathbf{x}^o)$ for all $k \in K$.*

Proof. Suppose, there exists $\mathbf{x}' \in Q$ which dominates \mathbf{x}^o . This means, $x'_i \geq x_i^o$ for all $i \in I$ with at least one inequality strict. Hence, $\bar{\theta}_k(\mathbf{x}') \geq \bar{\theta}_k(\mathbf{x}^o)$ for all $k \in K$ and $\bar{\theta}_{k_q}(\mathbf{x}') > \bar{\theta}_{k_q}(\mathbf{x}^o)$ which contradicts efficiency of \mathbf{x}^o within the restricted problem (20).

Suppose now that $\mathbf{x}' \in Q$ fairly dominates \mathbf{x}^o . Due to Theorem 3.1, this means that $\bar{\theta}_i(\mathbf{x}') \geq \bar{\theta}_i(\mathbf{x}^o)$ for all $i \in I$ with at least one inequality strict. Hence, $\bar{\theta}_k(\mathbf{x}') \geq \bar{\theta}_k(\mathbf{x}^o)$ for all $k \in K$ and any strict inequality would contradict efficiency of \mathbf{x}^o within the restricted problem (20). Thus, $\bar{\theta}_k(\mathbf{x}') = \bar{\theta}_k(\mathbf{x}^o)$ for all $k \in K$ which completes the proof. \square

It follows from Theorem 3.2 that while restricting the number of criteria in the multiple criteria model (14) we can essentially still expect reasonably fair efficient solution and only *unfairness* may be

related to the distribution of flows within classes of skipped criteria. In other words we have guaranteed some rough fairness while it can be possibly improved by redistribution of flows within the intervals $(\theta_{k_j}(\mathbf{x}), \theta_{k_{j+1}}(\mathbf{x}))$ for $j = 1, 2, \dots, q - 1$. Since the fairness preferences are usually very sensitive for the smallest flows, one may introduce a grid of criteria $1 = k_1 < k_2 < \dots < k_{q-1} < k_q = m$ which is dense for smaller indices while sparser for larger indices and expect solution offering some reasonable compromise between fairness and throughput maximization. In our computational analysis on the network with 132 elastic flows (Section 5) we have preselected 24 criteria including 12 the smallest flows. Note that any restricted model contains criteria $\bar{\theta}_1(\mathbf{x}) = \min_{i \in I} x_i$ and $\bar{\theta}_m(\mathbf{x}) = \sum_{i \in I} x_i$ among others. Therefore, it is more detailed than any bicriteria combination of max-min and throughput maximization.

4 Multiple criteria analysis

Following the equitable optimizations results discussed in the previous section, we may generate various fairly efficient bandwidth allocation patterns as efficient solutions of the multiple criteria problem:

$$\max (\eta_k)_{k \in K} \quad (21)$$

$$\text{subject to } \mathbf{x} \in Q$$

$$\eta_k = kt_k - \sum_{i \in I} d_{ik} \quad \text{for } k \in K \quad (22)$$

$$t_k - d_{ik} \leq x_i, \quad d_{ik} \geq 0 \quad \text{for } i \in I, k \in K \quad (23)$$

where $K \subseteq I$ and the attainable set Q is defined by constraints (1)–(6) and possibly (7) in the case of bandwidth modularity. Exactly, in the case of the complete multiple criteria model ($K = I$), according to Theorem 3.1, all fairly efficient allocations can be found as efficient solutions to (21)–(23) while in the case of restricted set of criteria $K \subset I$ some minor unfairness related to the distribution of flows within classes of skipped criteria may occur (Theorem 3.2).

The simplest way to model a large gamut of fairly efficient allocations may depend on the use some combinations of criteria $(\eta_k)_{k \in K}$. In particular, for the weighted sum with weights $w_k > 0$

$$\sum_{k \in K} w_k \eta_k = \sum_{k \in K} w_k \bar{\theta}_k(\mathbf{x}) = \sum_{i \in I} v_i \theta_i(\mathbf{x})$$

one apparently gets the so-called Ordered Weighted Averaging (OWA) [23] with weights $v_i = \sum_{k \in K: k \geq i} w_k$ ($i \in I$). If the weights v_i are strictly decreasing, i.e. in the case of full model ($K = I$), each optimal solution corresponding to the OWA maximization is a fair (fairly efficient) solution of (10) while the fairness among flows within classes of equal weights v_i (of skipped criteria) may be sometimes improved. Moreover, in the case of LP models, as the simplified network dimensioning (3)–(6), every fairly efficient allocation scheme can be identified as an OWA optimal solution with appropriate strictly monotonic weights [6]. Several decreasing sequences of weights provide us with various aggregations. Indeed, our earlier experience with application of the OWA criterion to the simplified problem of network dimensioning with elastic traffic [14] showed that we were able to generate easily allocations representing the classical fairness models. On the other hand, in order to find a larger variety of new compromise solutions we needed to incorporate some scaling techniques originating from the reference point methodology. Better controllability and the complete parameterization of nondominated solutions even for non-convex, discrete problems can be achieved with the direct use of the reference point methodology.

The reference point method was introduced by Wierzbicki [21] and later extended leading to efficient implementations of the so-called aspiration/reservation based decision support (ARBDS) approach with many successful applications [8, 22]. The ARBDS approach is an interactive technique allowing the DM to specify the requirements in terms of aspiration and reservation levels, i.e., by introducing acceptable and required values for several criteria. Depending on the specified aspiration and reservation levels, a special scalarizing achievement function is built which may be directly interpreted as expressing utility to be maximized. Maximization of the scalarizing achievement function generates an efficient solution to the multiple criteria problem. The solution is accepted by the DM or some modifications of the aspiration and reservation levels are introduced to continue the search for a better solution. The ARBDS approach provides a complete parameterization of the efficient set to multi-criteria optimization. Hence, when applying the ARBDS methodology to the ordered cumulated criteria in (14), one may generate all (fairly) equitably efficient solutions of the original resource allocation problem (10).

While building the scalarizing achievement function the following properties of the preference model are assumed. First of all, for any individual outcome η_k more is preferred to less (maximization). To meet this requirement the function must be strictly increasing with respect to each outcome. Second, a solution with all individual outcomes η_k satisfying the corresponding reservation levels is preferred to any solution with at least one individual outcome worse (smaller) than its reservation level. Next, provided that all the reservation levels are satisfied, a solution with all individual outcomes η_k equal to the corresponding aspiration levels is preferred to any solution with at least one individual outcome worse (smaller) than its aspiration level. That means, the scalarizing achievement function maximization must enforce reaching the reservation levels prior to further improving of criteria. In other words, the reservation levels represent some soft lower bounds on the maximized criteria. When all these lower bounds are satisfied, then the optimization process attempts to reach the aspiration levels.

The generic scalarizing achievement function takes the following form [21]:

$$\sigma(\eta) = \min_{k \in K} \{\sigma_k(\eta_k)\} + \varepsilon \sum_{k \in K} \sigma_k(\eta_k) \quad (24)$$

where ε is an arbitrary small positive number and σ_k , for $k \in K$, are the partial achievement functions measuring actual achievement of the individual outcome η_k with respect to the corresponding aspiration and reservation levels (η_k^a and η_k^r , respectively). Thus the scalarizing achievement function is, essentially, defined by the worst partial (individual) achievement but additionally regularized with the sum of all partial achievements. The regularization term is introduced only to guarantee the solution efficiency in the case when the maximization of the main term (the worst partial achievement) results in a non-unique optimal solution.

The partial achievement function σ_k can be interpreted as a measure of the DM's satisfaction with the current value (outcome) of the k -th criterion. It is a strictly increasing function of outcome η_k with value $\sigma_k = 1$ if $\eta_k = \eta_k^a$, and $\sigma_k = 0$ for $\eta_k = \eta_k^r$. Thus the partial achievement functions map the outcomes values onto a normalized scale of the DM's satisfaction. Various functions can be built meeting those requirements [22]. We use the piece wise linear partial achievement function introduced in [12]. It is given by

$$\sigma_k(\eta_k) = \begin{cases} \gamma(\eta_k - \eta_k^r)/(\eta_k^a - \eta_k^r), & \text{for } \eta_k \leq \eta_k^r \\ (\eta_k - \eta_k^r)/(\eta_k^a - \eta_k^r), & \text{for } \eta_k^r < \eta_k < \eta_k^a \\ \beta(\eta_k - \eta_k^a)/(\eta_k^a - \eta_k^r) + 1, & \text{for } \eta_k \geq \eta_k^a \end{cases} \quad (25)$$

where β and γ are arbitrarily defined parameters satisfying $0 < \beta < 1 < \gamma$. This partial achievement function is strictly increasing and concave which guarantees its LP computability with respect to

outcomes η_k .

Recall that in our model outcomes η_k represent cumulative ordered flows x_i , i.e. $\eta_k = \sum_{i=1}^k \theta_i(\mathbf{x})$. Hence, the reference vectors (aspiration and reservation) represent, in fact, some reference distributions of outcomes (flows). Moreover, due to the cumulation of outcomes, while considering equal flows ϕ as the reference (aspiration or reservation) distribution, one needs to set the corresponding levels as $\eta_k = k\phi$. Certainly, one may specify any desired reference distribution in terms of the ordered values of the flows (quantiles in the probability language) $\phi_1 \leq \phi_2 \leq \dots \leq \phi_m$ and cumulating them automatically get the reference values for the outcomes η_k representing the cumulated ordered flows. However, such rich modeling technique may be too complicated to control effectively the search for a compromise solution. Therefore, we rather consider to begin the search with a simplified approaches based on the reference flow distribution given as a linear sequence $\phi_k = \phi_1(1 + (k - 1)r)$ with the (relative) slope coefficient r thus leading to the cumulated reference levels increasing quadratically $\bar{\theta}_k(\phi) = \phi_1 k(2 + (k - 1)r)/2$. Although, special meaning of the the last (throughput) criterion should be rather operated independently from the others. Such an approach to control the the search for a compromise fair and efficient bandwidth allocation has been confirmed by the computational experiments as described in the following section.

5 Computational examples

The reference distribution approach described in preceding sections have been tested on a sample network dimensioning problem with elastic traffic. Recall that in the case of elastic traffic, the outcome of the network dimensioning procedure are the capacities of links in a given network, and that the flows will adapt to the bandwidth available on the links in the designed network. The input to a network dimensioning problem with elastic traffic consists of a network topology, of pairs of nodes that specify sources and destinations of flows, of sets of network paths that could be used for each flow, and of optional constraints on the capacities of links or on flow sizes. The user must also specify a budget for purchasing link capacity (B in (6)), prices of a unit of link capacity (possibly different for each link, c_e in (6)), and may specify module sizes and prices for a link. The given network topology may contain information about preinstalled link capacities (a_e in (3)): the budget is then spent on additional link capacities that extend the present capacity of a link.

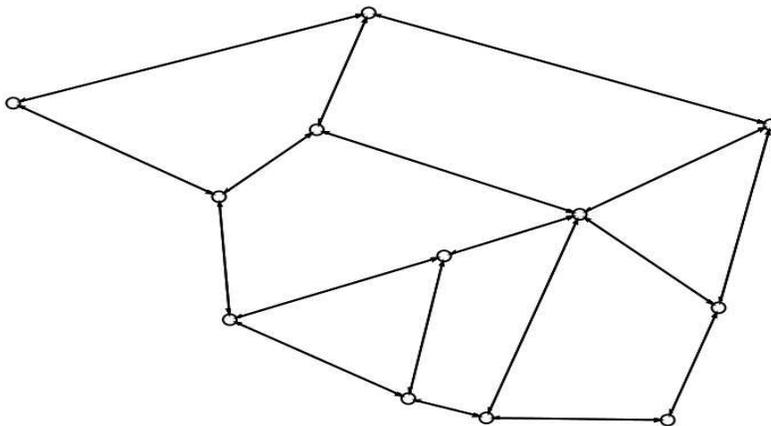


Figure 1: Sample network topology patterned after the backbone network of Polish ISP.

The network topology of the presented problem (Fig. 1) is patterned after the backbone network of a Polish ISP [14]. The network consists of 12 nodes and 18 links. Flows between any pair of different nodes have been considered (i.e., $144 - 12 = 132$ flows). For each flow, two alternative paths have been specified that could be used for transport. All information of a flow had to travel along one of the paths (nonbifurcation formulation (1)–(2)). All links have unit costs equal to one, and the budget for link bandwidth is $B = 1000$. Since all links have equal costs of one, path cost are equal to the path length (1, 2, 3 or 4 for the shortest paths in the example topology). All flows are unbounded. However, it is clear that due to the budget constraint no flow can exceed B .

In [14], a simplified LP model has been studied without additional constraints on link capacity, with a limitation that flows could only use the shortest path, and with equal link costs, since in such a case it was simple to understand the best choices with respect to fairness and overall throughput. However, for such a problem it is also simple to calculate the solution obtained by the two other methods used in literature for allocation problems with fairness objectives: max-min fairness and proportional fairness. Indeed, in [14] we have calculated these solutions and have shown the appropriate OWA aggregations allows us to obtain similar results. Additionally, using the OWA criterion, it was possible to obtain a spectrum of alternative solutions and to control the results using intuitive parameters. Here, we focus on two extensions of the problem studied in [14] that are too complex for a simple application of proportional fairness or max-min fairness. To apply either of these methods to the discussed problem extensions, it would be necessary to solve a nonlinear optimization problem or a sequence of MILP problems with changing constraints. The proposed problem modifications also make the studied models more practical and realistic.

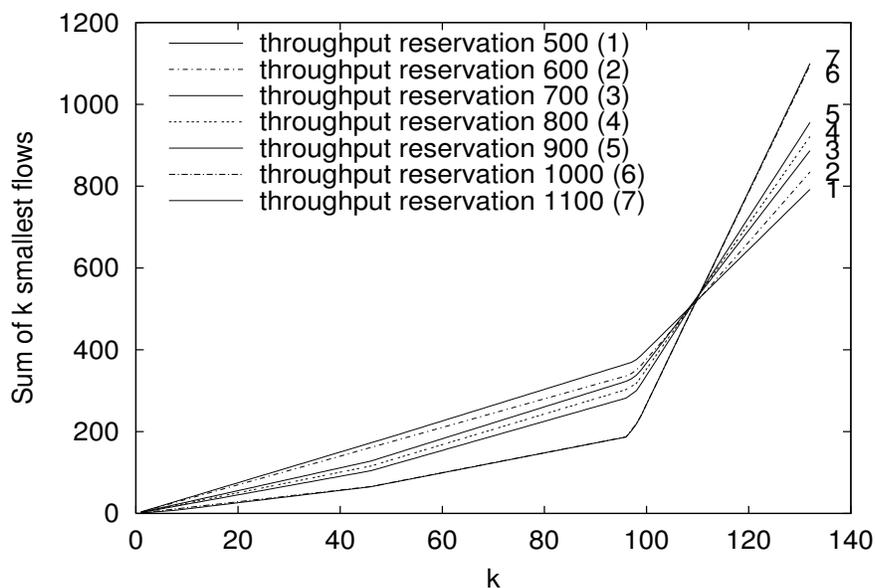


Figure 2: Flows distributions for varying throughput reservation with $r = 0.02$.

The first studied extension allowed flows to choose one of two paths for transport (1)–(2), added constraints that limited the capacity of certain links from above and added free link capacity for certain links (3). The intention behind the modification has been to model a situation when the network operator wishes to extend the capacity of an existing network. In this network, certain links cannot be upgraded beyond a certain values to the use of legacy technologies, due to prohibitive costs or administrative reasons (for instance, it may be cheap to use already installed fiber that has not been in use before, but it

may be prohibitively expensive to install additional fiber). The existence of free link capacity and of link capacity constraints may be the reason for choosing alternative paths for certain flows.

Additionally, a modular version of the original problem has been considered. In the second problem modification, flows were still limited to shortest paths, and no constraints on link capacities have been added. The size of a link capacity module was set to 5. For each link, an integer variable has been introduced (thus there are 18 integer variables in the modular version of the model). Modular link capacities are frequently encountered in networks, when it is simple to upgrade a link by installing an equipment module (like a new ATM card).

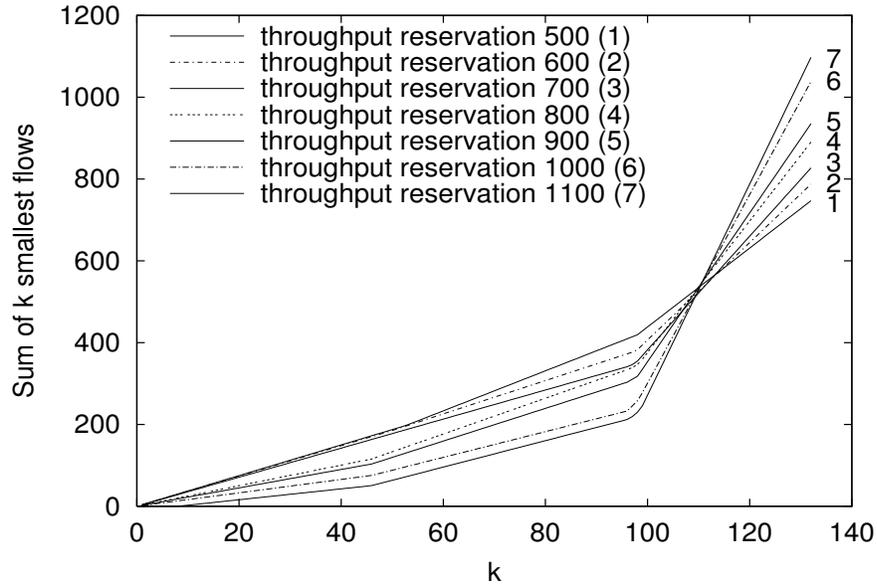


Figure 3: Flows distributions for varying throughput reservation with $r = 0.03$.

In [14], we have used all criteria η_k in the model, with the result that the linear program had to contain a large number of constraints (132^2). Here, we have limited the number of criteria η_k to 24, by choosing only the indices 1, 2, 3, ..., 9, 10, 11, 12, 18, 24, 30, 36, 48, 60, 72, ..., 120, 132 from the full set of all indices. As a result, the computation time has dropped from around one hour for each problem to the order of seconds. At the same time, the ability to control the outcomes using the reservation levels has not deteriorated; we were able to obtain similar results with the reduced set of criteria as with the full set.

For all model versions, the final input to the model consisted of the reservation and aspiration levels for the sums of ordered criteria. For simplicity, all aspiration levels were set close to the optimum values of the criteria, and only reservation levels were used to control the outcome flows. One of the most significant parameters was the reservation level for the sum of all criteria (the network throughput). This value denoted by η_m^r was selected (varying) separately from the other reservation levels. All the other reservation levels were formed following the linearly increasing sequence of the ordered values with slope (step) r and where the reservation level for minimal flow was taken $\phi_1 = 1$. Hence, for the final criteria $\eta_k = \bar{\theta}_k(\mathbf{x})$ representing the sums of ordered outcomes in model (17)–(19), the sequence of reservation levels increased quadratically (except from the last one). Thus, the three parameters have been used to define the reference distribution but we have managed to identify various fair and efficient allocation patterns by varying only two parameters: the reservation level η_m^r for the total throughput and the slope r for the linearly increasing sequence.

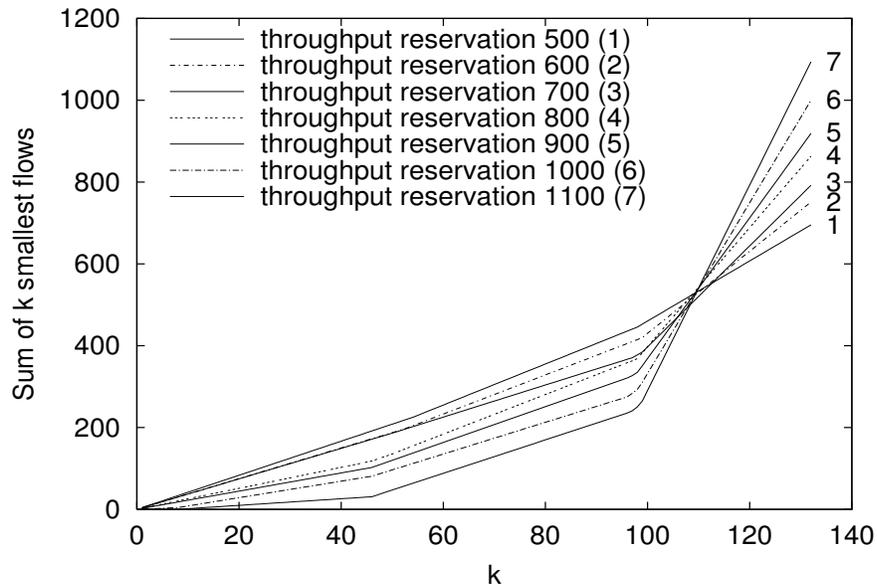


Figure 4: Flows distributions for varying throughput reservation with $r = 0.04$.

In the first experiment, we have used the first model extension that introduced alternative paths for flows, free link capacity and upper limits on capacity for certain links. For certain links, free link capacity was set to values from 5 to 20, and the upper limit on the capacity of certain links was set to 20. Due to the presence of free link capacity and upper limits on link capacity, the MILP solver found solutions where certain flows had to use alternative paths rather than shortest paths. These flows were more expensive than other flows that were allowed to use their shortest paths. Recall that we have used a single-path formulation, meaning that the entire flow had to be switched to the alternative path. Flows could not be split, which is consistent with several traffic engineering technologies used today.

In the experiment, the reservation level η_m^r and the slope r have been used to search for compromise solutions that traded off fairness against efficiency. The throughput reservation has been varied from 500 to 1100. As η_m^r increases, the cheaper flows receive more throughput at the expense of more expensive (longer) flows. For values of η_m^r above 1100, some flows were starved, and therefore these outcomes were not considered further.

The linear increase of the other reservation levels was varied as well. The parameter r could have values of: 0.02, 0.03 and 0.04. The results of the experiment are shown in Fig. 2–4 with the corresponding absolute Lorenz curves [7]. The figures present plots of cumulated ordered flows $\bar{\theta}_k(\mathbf{x})$ versus number k (rank of a flow in ordering according to flow throughput) which means that the normalizing factor $1/m = 1/132$ has been ignored (for both the axes). The total network throughput is represented in the figures by the altitude of the right end of the curve ($\bar{\theta}_{132}(\mathbf{x})$). A perfectly equal distributions of flows would be graphically represented by an ascending line of constant slope.

Note that under moderate throughput requirements, as r increases, the medium flows gain at the expense of the larger ones thus enforcing more equal distribution of flows (one may observe flattening of the curves). On the other hand, with higher throughput reservations the larger flows are protected by this requirement and increase of r causes that the medium flows gain at the expense of the smallest flows (one may observe convexification of the curves). For values of r higher than 0.04, the increase of the throughput reservation resulted in flow starvation.

Observe from Fig. 4 that for $\eta_m^r = 1100$ (and for some other values of η_m^r), the boundary between the largest flows (part of the Lorenz curve with the highest slope) and the second-largest flows is not sharp. The change of slope is gradual, resembling a round knee. The reason for this is the presence of three flows that should receive the same amount of throughput as the largest flows, since they are all transported on paths of the length of 1, but cannot due to the presence of upper constraints on link capacities. These flows receive as much as they can, but some capacity is left for other flows that must travel on the same constrained link. Here the model violates fairness in the attempt to get a higher total network throughput.

Also, note on the same figure that the boundary between the smallest flows for $\eta_m^r = 500$ and for $\eta_m^r = 1100$ is not in the same position. The reason for this is once again the upper constraint on link capacities. For $\eta_m^r = 500$, there are 8 flows that should be in the middle group of flows but cannot, since flows in the middle group receive so much throughput that the constraints on link capacity would be violated. Therefore, these flows are downgraded to the group of smallest flows and receive the same amount of throughput as the smallest flows – here the model preserves fairness.

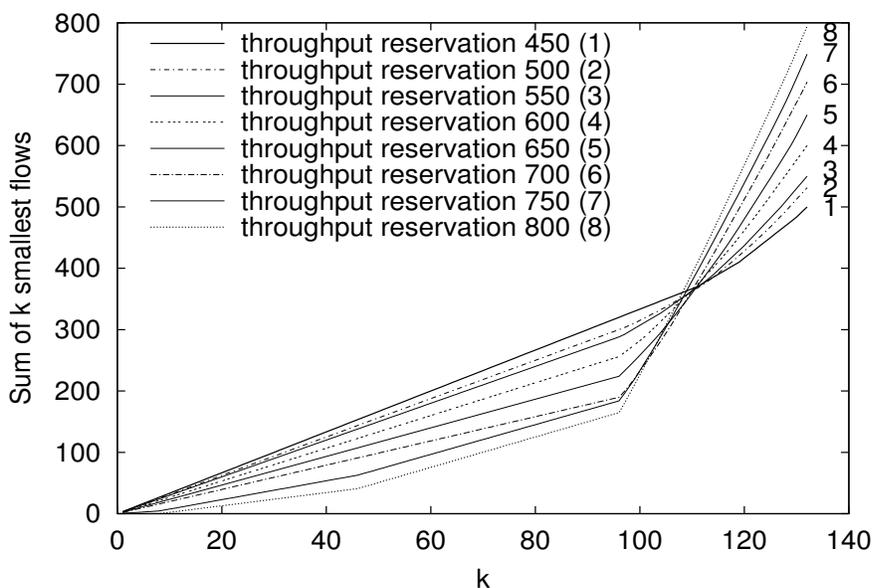


Figure 5: Varying throughput reservation with $r = 0.02$ for the model with bandwidth modularity.

Note that the throughput reservation was effectively used to find outcomes with the desired network throughput. On the other hand, especially for large throughput reservations, the optimization procedure automatically found outcomes that divided flows into four categories according to their path costs. This shows that the presented methodology is cost-aware, and that it is possible to guarantee fairness to all flows with the same path cost (if link capacity constraints do not interfere). For the lowest throughput reservation of $\eta_m^r = 500$ and $r = 0.04$, the outcome was close to a perfectly fair distribution. Using the methodology described in this paper, the user can choose from a large number of different outcomes and control the trade-off between fairness and efficiency.

For the second experiment we used a slightly different dimensioning problem specification. Namely, we included the modular link capacities (7) into the model while eliminating the routing decisions by restriction of all service flows to the corresponding shortest paths. Thus, the model was still in the class of MILP but with different discrete structure. We have repeated a search for compromise solutions

using similar preference parameter configurations as in the first experiment, although taking into account the constraints on modular bandwidth, the throughput reservation has been varied from 450 to 800. The resulting flows distributions for the reservation slope parameter $r = 0.02$ are presented in Fig. 5. Predictably, the introduction of modular link capacities makes it more difficult to find fair solutions. The outcome for $\eta_m^r = 450$ is close to a perfectly even (fair) distribution, although the right end of the curve turns slightly upward. This indicates that the excess capacities of modules were used by the cheapest flows, leading to a higher network throughput than in the case of a problem without modular link capacities. On the other hand the flows with the cheapest paths were not equal for some outcomes.

Overall, the two experiments on the sample network topology demonstrated the versatility of the described methodology for equitable optimization. The use of reservation levels, controlled by a small number of simple parameters, allowed us to search for solutions best fitted to various possible preferences of a network designer. Using appropriate reference point based procedure, one should be able easily find a satisfactory fair and efficient allocation pattern in a few interactive steps.

6 Concluding remarks

A central issue in networking is how to allocate bandwidth to flows efficiently and fairly. The so-called Max-Min Fairness is widely used to meet these goals. Allocating the resources to optimize the worst performances may cause a large worsening of the overall (mean) performances. Therefore, several other fair allocation schemes are searched and analyzed. Our earlier computational experiments with application of the OWA criterion to the (simplified) LP problem of network dimensioning with elastic traffic [14] showed that we were able to generate easily allocations representing the classical fairness models. On the other hand, in order to find a larger variety of new compromise solutions we needed to incorporate some scaling techniques originated from the reference point methodology. Actually it is a common flaw of the weighting approaches that they provide poor controllability of the preference modeling process and in the case of multiple criteria problems with discrete (or more general nonconvex) feasible sets, they may fail to identify several compromise efficient solutions [19].

In standard multiple criteria optimization, good controllability and the complete parameterization of nondominated solutions can be achieved with the direct use of the reference point methodology. While looking for fairly efficient bandwidth allocation the reference point methodology can be applied to the cumulated ordered outcomes. Our initial experiments with such an approach to the problem of network dimensioning with elastic traffic have confirmed the theoretical advantages of the method. We were able to generate easily various (compromise) fair solutions for both continuous and modular problems despite the search for fairly efficient compromise solutions was controlled by only two parameters. One of these parameters was a reservation level for the network throughput. The second parameter allowed the network designer to control the difference in throughputs of cheaper and more expensive flows. Still, flows with the same cost were always treated fairly. Moreover, the obtained solutions divided flows into categories determined by flow cost. For modular solutions, the cheapest flows consumed the excess link capacity. These characteristics demonstrate that the model is cost-aware and fulfills the axioms of equitable optimization.

Appendix

We have also tested an alternative scheme of the preference modeling within our reference point method implementation. Namely, we analyzed the initial scheme (see Section 4) based on the reference flow distribution given as a linear sequence $\phi_k = \phi_1(1 + (k - 1)r)$ with the (relative) slope coefficient r thus leading to the cumulated reference levels increasing quadratically $\bar{\theta}_k(\phi) = \phi_1 k(2 + (k - 1)r)/2$ is strictly implemented. The sequence was applied to construct all the reservation levels including η_1^r for the minimum flow and η_m^r for the network throughput. Although the value of η_m^r , due to the represented throughput criterion, had to be selected (varying) directly. Therefore, all the other reservation levels were formed according to the linearly increasing sequence of the ordered values with slope (step) r where the reservation level for the minimal flow ϕ_1 had allocated a value guaranteeing that $\eta_m^r = \phi_1 m(2 + (m - 1)r)/2$. Thus, the two parameters have been used to define the reference distribution: the reservation level η_m^r for the total throughput and the slope r for the linearly increasing sequence but (opposite to the scheme from Section 5) ϕ_1 has not been fixed.

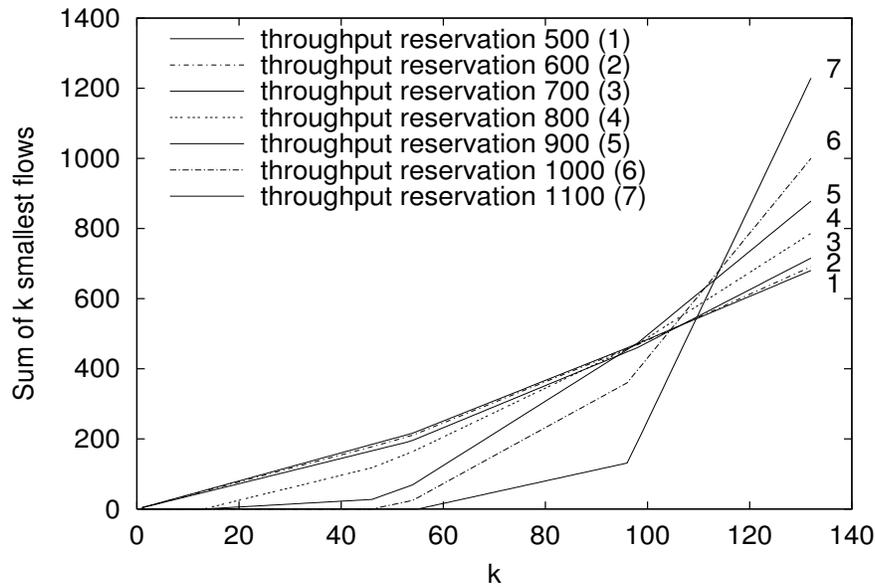


Figure 6: Results for varying throughput reservation with $r = 0.02$ defining all other reservations.

We have applied this preference model to the first bandwidth allocation problem consisted of the single paths requirements, free link capacity and upper limits on capacity for certain links. The results of the experiment with $r = 0.02$ and varying η_m^r are shown in Fig. 6 with the corresponding absolute Lorenz curves. As η_m^r increases, the cheaper flows receive more throughput at the expense of more expensive (longer) flows. It turned out that except from relatively minor throughput requirements (values 500 to 700), increasing values of η_m^r introduced significant inequity among flows and numerous flows were starved. Similar solutions appeared for various values of r . Therefore, we have abandoned such a two parameter control scheme and we have decided that the throughput criterion should be rather operated independently from the others. Such an approach to control the search for a compromise fair and efficient bandwidth allocation has been confirmed by the computational experiments as described in Section 5.

References

- [1] Bertsekas D, Gallager R. Data Networks. Englewood Cliffs: Prentice-Hall, 1987.
- [2] Bonald T, Massoulié L. Impact of fairness on Internet performance. Proceedings of ACM Sigmetrics, June 2001, 82–91.
- [3] Denda R, Banchs A, Effelsberg W. The fairness challenge in computer networks. Lect Notes in Comp Sci 2000; 1922: 208–220.
- [4] Jaffe J. Bottleneck flow control. IEEE Trans. on Communications 1980; 7:207–237.
- [5] Kelly F, Mauloo A, Tan D. Rate control for communication networks: shadow prices, proportional fairness and stability. J Oper Res Soc 1997; 49:206–217.
- [6] Kostreva MM, Ogryczak W. Linear optimization with multiple equitable criteria. RAIRO Oper Res 1999; 33:275–297.
- [7] Kostreva MM, Ogryczak W, Wierzbicki A. Equitable aggregations and multiple criteria analysis, European J Opnl Res 2004; 158:362–367.
- [8] Lewandowski A, Wierzbicki AP. Aspiration Based Decision Support Systems — Theory, Software and Applications. Berlin: Springer, 1989.
- [9] Luss H. On equitable resource allocation problems: a lexicographic minimax approach. Oper Res 1999; 47:361–378.
- [10] Marshall AW, Olkin I. Inequalities: Theory of Majorization and Its Applications. New York: Academic Press, 1979.
- [11] Mo J, Walrand J. Fair end-to-end window-based congestion control. IEEE/ACM Trans on Networking 2000; 8:556–567.
- [12] Ogryczak W. Linear and Discrete Optimization with Multiple Criteria: Preference Models and Applications to Decision Support (in Polish). Warsaw: Warsaw Univ Press, 1997.
- [13] Ogryczak W, Śliwiński, T. On equitable approaches to resource allocation problems: the conditional minimax solution. J Telecomm and Info Tech 2002; 3:40–48.
- [14] Ogryczak W, Śliwiński T, Wierzbicki A. Fair resource allocation schemes and network dimensioning problems. J Telecomm and Info Tech 2003; 3:34–42.
- [15] Ogryczak W, Tamir A. Minimizing the sum of the k largest functions in linear time. Information Proc. Letters 2003; 85:117–122.
- [16] Pióro M, Medhi D. Routing, Flow and Capacity Design in Communication and Computer Networks. San Francisco: Morgan Kaufmann Publ, 2004.
- [17] Rawls J. The Theory of Justice. Cambridge: Harvard University Press, 1971.
- [18] Steinhaus H. Sur la division pragmatique. Econometrica 1949; 17:315–319.

- [19] Steuer RE. Multiple Criteria Optimization: Theory, Computation & Applications. New York: Wiley, 1986.
- [20] Tang A, Wang J, Low SH. Is fair allocation always inefficient. IEEE INFOCOM 2004.
- [21] Wierzbicki AP. A mathematical basis for satisficing decision making. Math Modelling 1982, 3:391–405.
- [22] Wierzbicki AP, Makowski M, Wessels J. Model Based Decision Support Methodology with Environmental Applications. Dordrecht: Kluwer, 2000.
- [23] Yager RR. On ordered weighted averaging aggregation operators in multicriteria decision making. IEEE Trans Systems, Man and Cybernetics 1988; 18:183–190.