Paper

On equitable approaches to resource allocation problems: the conditional minimax solutions

Włodzimierz Ogryczak and Tomasz Śliwiński

Abstract — In this paper we introduce and analyze a solution concept of the conditional minimax as a generalization of the minimax solution concept extended to take into account the number of services (the portion of demand) related to the worst performances. Namely, for a specified portion of demand we take into account the corresponding portion of the maximum results and we consider their average as the worst conditional mean to be minimized. We show that, similar to the standard minimax approach, the minimization of the worst conditional mean can be defined by a linear objective and a number of auxiliary linear inequalities. We report some results of initial computational experience with the new solution concept.

Keywords* — *telecommunication networks, resource allocation, equity, min-max.

1. Introduction

Resource allocation problems are concerned with the allocation of limited resources among competing activities so as to achieve the best overall performances. In this paper, we focus on approaches that, while allocating resources, attempt to provide an equal treatment of all the competing activities [8]. The problems of efficient and equitable resource allocation arise in various systems which serve many users, like in telecommunication systems among others. Telecommunication networks are expected to satisfy the increasing demand for traditional services as well as to accommodate multimedia services. Hence, it becomes critical to allocate network resources, such as available bandwidth, so as to provide high level performance of all services at numerous destination nodes. The performance can be measured in terms of expected delays to be equitably minimized for all service demands.

The generic resource allocation problem may be stated as follows. Each activity is measured by an individual performance function that depends on the corresponding resource level assigned to that activity. A smaller function value is considered better, like the performance measured in terms of expected delays. Models with an (aggregated) objective function that minimizes the mean (or simply the sum) of individual performances are widely used to formulate resource allocation problems, thus defining the so-called minisum solution concept. This solution concept is primarily concerned with the overall system efficiency. As based on averaging, it often provides solution where low demand services are discriminated in terms of delays. An alternative approach depends on the so-called minimax solution concept, where the worst performance (maximum delay) is minimized. The minimax approach is consistent with Rawlsian [11] theory of justice, especially when additionally regularized with the lexicographic order [9]. On the other hand, allocating the resources to optimize the worst performances may cause a large worsening of the overall (mean) performances.

In this paper we introduce and analyze an alternative compromise solution concept of the conditional minimax. It is a parametric generalization of the minimax solution concept taking into account the number of services (the portion of demand) related to the worst performances. Namely, for a specified tolerance level (number of services k or portion of demand β) we take into account the entire group of the k (β portion) maximum results and we consider their average as the worst conditional mean to be minimized. According to this definition the solution concept is based on averaging restricted to the group of the worst results. We show that, similar to the standard minimax approach, the minimization of the worst conditional mean can be defined by a linear objective and a number of auxiliary linear inequalities.

Resource allocation models may be used to help to solve two major types of telecommunication problems emerging with exploding demand on multimedia services [2]. The first type of problems is related to decision support for designing robust and cost-effective fiber-optic networks [3]. The other field is traffic engineering which represents the ability to optimize the use of network resources only by means of efficient routing decisions [4]. In other words, while the first group of problems deals with the network engineering being related to the physical design of the network, the second group is rather related to the software design. The proposed solution approach is general enough to be applicable for both types of problems. However, we demonstrate it on straightforward problems related to the traffic engineering.

The paper is organized as follows. In the next section we introduce our generic resource allocation model and we show how it can be used to express several traffic engineering problems. In Section 3 the solution concept of the conditional minimax is formally introduced and it is shown that, similar to the standard minimax approach, the solution can be defined by a linear objective and a number of auxiliary linear inequalities. In Section 4 we report some results of our initial computational experience with the new solution concept.

2. The model

The generic resource allocation problem that we consider may be stated as follows. There is given a set of m services. There is also given a set Q of allocation patterns (allocation decisions). For each service i a function $f_i(\mathbf{x})$ of the allocation pattern $\mathbf{x}$ has been defined. This function, called the individual objective function, measures the outcome (effect) $y_i = f_i(\mathbf{x})$ of the allocation pattern for service i. In applications, we consider, an outcome usually expresses the delay. However, we emphasize to the reader that we do not restrict our considerations to the case of outcomes measured as delays. They can be measured (modeled) as service time, service costs as well as in a more subjective way. In typical formulations a smaller value of the outcome (delay) means a better effect (higher service quality or client satisfaction). Otherwise, the outcomes can be replaced with their complements to some large number. Therefore, without loss of generality, we can assume that each individual outcome y_i is to be minimized which results in a multiple criteria minimization model.

The simplest services structure forms the uniform problem where each service represents a single unit. Usually, one is interested in putting into allocation model some additional demand weights $w_i > 0$ to represent the amount of demand for the specific service. Integer weights can be interpreted as numbers of unweighted identical services to be repeated independently. For initial theoretical considerations we will assume that the problem is transformed (disaggregated) to the uniform one (that means all the demand weights are equal to 1). Note that such a disaggregation is possible for integer as well as rational demand weights, but it usually dramatically increases the problem size. Therefore, we consider solution concepts which can be applied directly to the weighted problem. For this purpose we will use the normalized demand weights

$$\bar{w}_i = w_i / \sum_{j=1}^{m} w_j \quad \text{for } i = 1,2,\dots,m \tag{1}$$

rather than the original quantities w_i. Note that, in the case of uniform problem (all $w_i = 1$), all the normalized weights are given as $\bar{w}_i = 1/m$.

Telecommunication problems deal with routing of the data traffic in an existing network or with designing the network expansions to accommodate the traffic. Both type of problems require the allocation of network resources (capacities or potential capacities). Let us consider a connected network consisted of a node set N to represent various locations. Directed links $(j,k) \in L \subset N \times N$ are attributed by the bandwidth/capacity coefficients b_{jk} and the delay/distance/cost coefficients c_{jk}. Further, we consider a set $I = \{1,2,\dots,m\}$ of m services. Each service is related to some data traffic between two network nodes. Thus, the service is described by a directed pair of nodes $\big(s(i), d(i)\big)$ representing the source and the destination of the data traffic, respectively. The amount of the data traffic related to service i is described by the demand weight w_i. The latter may be skipped in the case of uniform problem where all the services generate the same amount of data traffic ($w_i = 1$ for all i).

Within a telecommunication network the data traffic is generated by a huge number of nodes exchanging data. In such a network, a relatively small subset $H \subset N$ of nodes are chosen to serve as hubs which can be used as intermediate switching points [1, 6]. Given a set of hubs, data traffic generated by a service is sent from the source node to a hub first. It can be then sent along communications link between hubs, and finally reach the destination node along a link from a hub. The hub-based network organization allows the data traffic to be consolidated on the inter-hub links.

While taking into account the hub-based network structure, the main decisions to be made for the services organization can be described with the assignment of a directed pair of hubs $\big(h'(i), h''(i)\big)$ to each service i. The data traffic for service i is then implemented by sending from the source $s(i)$ to the hub $h'(i)$ first, the use of the inter-hub connection from $h'(i)$ to $h''(i)$ next, and the final sending from $h''(i)$ to the service destination $d(i)$. The delay/distance of such a data path is usually assumed to be defined as the sum of several link delays $c_{s(i),h'(i)} + c_{h'(i),h''(i)} + c_{h''(i),d(i)}$. Note that a single hub can be used is some cases $\big(h'(i) = h''(i)\big)$ which may require a definition of the corresponding dummy inter-hub links.

In the case of the demand weights for various services there is no justification for a strict assignment of a single path to the specific service since several units may be sent along different paths. Therefore, the main decisions may be modeled with variables $x_{ijk}\big(i \in I;\ j,k \in H\big)$ expressing the amount of data traffic related to service i routed via hubs h_j and h_k. To meet the problem requirements, the decision variables x_{ijk} have to satisfy the following constraints:

$$\sum_{j\in H}\sum_{k\in H} x_{ijk} = w_i \qquad \text{for } i \in I, \tag{2}$$

$$\sum_{i\in I} x_{ijk} \le b_{jk} \qquad \text{for } j,k \in H, \tag{3}$$

$$x_{ijk} \ge 0 \qquad \text{for } i \in I;\ j,k \in H, \tag{4}$$

where Eqs. (2) guarantee the routing of whole service demands while inequalities (3) keep the data traffic within the capacity limits. Note that taking into account the hub-based network specificity we have considered the capacity constraints only for the inter-hub links.

The unit performance measure (delay) of the service i may be expressed with the following linear function:

$$f_i(\mathbf{x}) = \frac{1}{w_i} \sum_{j,k\in H} \big[c_{s(i),j} + c_{jk} + c_{k,d(i)}\big] x_{ijk} \quad \text{for} \quad i \in I. \tag{5}$$

Hence, all the functions $f_i(\mathbf{x})$ need to be minimized. The typical problems involving routing decisions are considered as dynamic and stochastic. Nevertheless, one may analyze a straightforward static allocation problem related to traffic engineering (routing) decisions within a telecommunication (or transportation) network. Such a problem depends, simply, on minimization of criteria (5) subject to constraints (2)–(4).

Similar model may be considered for the inter-hub bandwidth allocation problem related to the network design issues. Namely, one may minimize the same criteria (5) and the same constraints (2)–(4), but the bandwidth b_{jk} in constraints (3) need to be considered decision variables rather than data parameters. Again, it is a straightforward network design model but its analysis may be useful at some initial phases of the design process.

In the above model we allow the services to be partitioned in various portions of the demand and implemented with possibly different routing. We believe that is acceptable for most applications related to data transfer as the standard data package is relatively extremely small when comparing to the total amount of demand. Moreover, new routing protocols developed for the Internet services, like the multi-protocol label switching, allow much flexibility in the traffic engineering solutions [4]. Nevertheless, the problem (2)–(5) may be adapted to the requirement of a single route assigned to each service, if necessary. For this purpose, one needs to introduce binary decision variables x_{ijk} equal 1 when the inter-hub link (h_j, h_k) is used to implement service i, and 0 otherwise. The constraints take then the following form:

$$\sum_{j\in H}\sum_{k\in H} x_{ijk} = 1 \qquad \text{for} \quad i \in I,$$

$$\sum_{i\in I} w_i x_{ijk} \le b_{jk} \qquad \text{for} \quad j,k \in H$$

and the resulting model is very close to the location problems [6, 10].

In problem (2)–(5) we have considered all the hubs as directly connected by the corresponding inter-hub links. In telecommunication networks hubs are rather organized in some network structure (architecture) which causes the existence of some interactions (common bandwidth limits) between various inter-hub connections representing rather paths (routes) than direct links. The modern telecommunication networks heavily use the architecture of a collection of bidirectional rings (as in SONET standard [3]). Below we specify in details such an allocation model where the hubs are arranged in a cycle and the traffic engineering problem needs to take into account the bidirectional ring-loading issues. This type of models we will use in Section 5 to demonstrate some computational results.

Let us consider again a connected network consisted of a node set N directed links $(j,k) \in L \subset N \times N$ which are attributed by the bandwidth/capacity coefficients b_{jk} and the delay/distance/cost coefficients c_{jk}. A set $I = \{1,2,\ldots,m\}$ of m services is considered. Each service is related to a directed pair of nodes $(s(i),d(i))$ (the source and the destination of the data traffic), and it requires the amount w_i of the data traffic (demand weight w_i). A relatively small subset $H \subset N$ of p nodes are chosen to serve as hubs. Hubs $h_1, h_2, \ldots, h_{p-1}, h_p$ are arranged clockwise in a cycle (ring). That means, there are p clockwise directed inter-hub links: $(h_1,h_2),(h_2,h_3), \ldots, (h_{p-1},h_p)$, (h_p,h_1), and p counterclockwise directed inter-hub links: $(h_p,h_{p-1}),(h_{p-1},h_{p-2}), \ldots, (h_2,h_1)$, (h_1,h_p).

The main decisions may be modeled with variables x'_{ijk} and x''_{ijk} $(i \in I;\ j,k \in H)$ expressing the amount of data traffic related to service i routed via hubs h_j and h_k using clockwise or counterclockwise connection, respectively. To meet the service demand requirements, the decision variables have to satisfy the following constraints:

$$\sum_{j\in H}\sum_{k\in H}(x'_{ijk} + x''_{ijk}) = w_i \qquad \text{for } i \in I\,, \tag{6}$$

$$x'_{ijk}, x''_{ijk} \ge 0 \qquad \text{for } i \in I;\ j,k \in H\,. \tag{7}$$

Recall that there is a piece of data traffic which passes trough a single hub not generating the ring traffic either clockwise or counterclockwise. Namely, $x'_{ijj} + x''_{ijj}$ for $j \in H$ is the amount of such traffic and it could be represented by a single variable but we have accepted the redundancy to keep the constraints (6) simpler.

To analyze the bandwidth (links capacity) allocation one needs to accumulate the traffic load of specific links in the ring. Let $(l_1,l_2) \in C$ denote a clockwise link in the ring, i.e. $l_2 = l_1 + 1$ for $l_1 = 1,\ldots,p-1$ or $l_2 = 1$ for $l_1 = p$. The link is loaded with clockwise traffic from hub h_j to hub h_k for $j = 1,\ldots,l_2-1$ and $k = l_2,\ldots,p$ or $k = 1,\ldots,j-1$ as well as (if $l_2 \le p-1$) for $j = l_2+1,\ldots,p$ and $k = l_2,\ldots,j-1$. Hence, the clockwise traffic of all the services generates the following (clockwise) link load

$$z'_{l_1,l_2} = \sum_{i\in I}\left[\sum_{j=1}^{l_2-1}\Big(\sum_{k=l_2}^{p} x'_{ijk} + \sum_{k=1}^{j-1} x'_{ijk}\Big) + \sum_{j=l_2+1}^{p}\sum_{k=l_2}^{j-1} x'_{ijk}\right] \tag{8}$$

for each $(l_1,l_2) \in C$. By symmetry, the counterclockwise traffic of all the services generates the (counterclockwise) link load

$$z''_{l_1,l_2} = \sum_{i\in I}\left[\sum_{k=1}^{l_2-1}\Big(\sum_{j=l_2}^{p} x''_{ijk} + \sum_{j=1}^{k-1} x''_{ijk}\Big) + \sum_{k=l_2+1}^{p}\sum_{j=l_2}^{k-1} x''_{ijk}\right] \tag{9}$$

for each $(l_1,l_2) \in C$. Note that z''_{l_1,l_2} denotes, in fact, the load of directed counterclockwise link (l_2,l_1). With commonly considered bidirectional capacity (bandwidth) limits the link loads must satisfy the constraints

$$z'_{l_1,l_2} + z''_{l_1,l_2} \le b_{l_1,l_2} \text{for } (l_1,l_2) \in C\,. \tag{10}$$

In the case of independently considered separate single-directional capacity limits, the latter needs to be replaced with constraints

$$z'_{l_1,l_2} \le b'_{l_1,l_2} \quad \text{and} \quad z''_{l_1,l_2} \le b''_{l_1,l_2} \quad \text{for} \quad (l_1,l_2) \in C\,.$$

The unit performance measure (delay) for the service i is expressed with the following linear function:

$$f_i(\mathbf{x}) = \frac{1}{w_i}\sum_{j,k\in H}(c_{s(i),j} + d'_{jk} + c_{k,d(i)})x'_{ijk} + \\ + \frac{1}{w_i}\sum_{j,k\in H}(c_{s(i),j} + d''_{jk} + c_{k,d(i)})x''_{ijk}, \qquad (11)$$

where d'_{jk} and d''_{jk} denote the delays along the clockwise and counterclockwise, respectively, paths from h_j to h_k in the ring C. For instance, in the case of $1 \le j < k \le p$ one gets $d'_{jk} = c_{j,j+1} + \ldots + c_{k-1,k}$. Certainly, all the functions $f_i(\mathbf{x})$ need to be minimized. Hence, a simple traffic engineering problem with bidirectional ring loading issues can be considered as multiple criteria minimization of (11) subject to constraints (6)–(10).

The problem (6)–(11) may be adapted to the requirement of a single inter-hub route assigned to each service, if necessary. Let us assume that the data traffic related to service i and routed via hubs h_j and h_k has to use either clockwise or counterclockwise connection without any splitting. This requirement can be modeled by introducing binary decision variables r_{ijk} equal 1 when the clockwise connection from h_j to h_k is used to implement service i, and 0 for the counterclockwise connection. The model needs to be extended then with the constraints of the following form:

$$x'_{ijk} \le w_i r_{ijk} \quad \text{and} \quad x'_{ijk} \le w_i(1 - r_{ijk}) \quad \text{for} \quad i \in I,\ j,k \in H.$$

One may also formulate a network design problem where quantities b_{l_1,l_2} for $(l_1,l_2) \in C$ are considered as a set of multiple criteria to be minimized subject to constraints (6)–(10) with possible upper limits on service delays $f_i(\mathbf{x})$.

3. The solution concept

Assuming that the generic allocation problem has been disaggregated to the unweighted form (all $w_i = 1$), it may be stated as the following multiple criteria minimization problem:

$$\min\ \{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in Q\}, \qquad (12)$$

where $\mathbf{f} = (f_1, \ldots, f_m)$ is a vector of the individual objective functions which measure the outcome (effect) $y_i = f_i(\mathbf{x})$ of the allocation pattern $\mathbf{x}$ for service i.

We do not assume any special form of the feasible set while introducing the solution concepts. We rather allow the feasible set to be a general, possibly discrete (nonconvex), set. Similarly, we do not assume any special form of the individual objective functions nor their special properties (like convexity). Therefore, the solution concepts may be applied to various allocation problems. Nevertheless, the solution concepts, we consider, are implementable by a linear objective and a number of auxiliary linear inequalities. Thus the solution concepts preserve a possible structure (LP or convexity) of the allocation problem under analysis.

Most classical allocation studies focus on the minimization of the mean (or total) outcome or the minimization of the maximum (the worst) outcome. Both the corresponding solution concepts are well defined for aggregated allocation models using demand weights $w_i > 0$. Exactly, for the weighted allocation problem, the *minisum* solution concept is defined by the minimization of the objective function expressing the *mean* (average) outcome

$$\mu(\mathbf{y}) = \sum_{i=1}^{m} \bar{w}_i y_i$$

but it is also equivalent to the minimization of the total outcome $\sum_{i=1}^{m} w_i y_i$. The *minimax* solution concept is defined by the minimization of the objective function representing the *maximum* (worst) outcome

$$M(\mathbf{y}) = \max_{i=1,\ldots,m} y_i$$

and it is not affected by the demand weights at all. Both the classical solution concepts are represented with simple aggregation of multiple criteria model (12). Namely, the minisum approach simply use the weighted sum of criteria

$$\min\Big\{\sum_{i=1}^{m} \bar{w}_i f_i(\mathbf{x}) : \mathbf{x} \in Q\Big\} \qquad (13)$$

while the minimax approach results in a problem

$$\min\big\{t : \mathbf{x} \in Q;\ t \ge f_i(\mathbf{x}) \quad \text{for } i = 1,2,\ldots,m\big\} \qquad (14)$$

with only one auxiliary variable t and m inequalities to define it.

Since the minisum approach is based on averaging, it often provides solutions where low demand services related to remote destinations are discriminated in terms of delays. On the other hand, allocating the resources to optimize the worst case may cause a large increase in the total delays thus generating a substantial loss in the overall system efficiency. This has led to a search for some compromise solution concept.

A natural generalization of the maximum (worst) outcome $M(\mathbf{y})$ is the worst conditional mean defined as the mean within the specified tolerance level (amount) of the worst outcomes. For the simplest case of the unweighted allocation problem (12), one may simply define the worst conditional mean $M_{\frac{k}{m}}(\mathbf{y})$ as the mean outcome for the k worst-off services (or rather k/m portion of the worst services). This can be mathematically formalized as follows. First, we introduce the ordering map $\Theta : R^m \to R^m$ such that $\Theta(\mathbf{y}) = (\theta_1(\mathbf{y}), \theta_2(\mathbf{y}), \ldots, \theta_m(\mathbf{y}))$, where $\theta_1(\mathbf{y}) \ge \theta_2(\mathbf{y}) \ge \cdots \ge \theta_m(\mathbf{y})$ and there exists a permutation τ of set I such that $\theta_i(\mathbf{y}) = y_{\tau(i)}$ for $i = 1,2,\ldots,m$. The use of ordered outcome vectors $\Theta(\mathbf{y})$ allows us to focus on distributions of outcomes impartially. Next, the linear cumulative map is applied to ordered outcome vectors to get $\bar{\Theta}(\mathbf{y}) = (\bar{\theta}_1(\mathbf{y}), \bar{\theta}_2(\mathbf{y}), \ldots, \bar{\theta}_m(\mathbf{y}))$ defined as

$$\bar{\theta}_k(\mathbf{y}) = \sum_{i=1}^{k} \theta_i(\mathbf{y}), \quad \text{for } k = 1,2,\ldots,m. \qquad (15)$$

The coefficients of vector $\overline{\Theta}(\mathbf{y})$ express, respectively: the largest outcome, the total of the two largest outcomes, the total of the three largest outcomes, etc. Hence, the *worst k/m–conditional mean* $M_{\frac{k}{m}}(\mathbf{y})$ is given as

$$M_{\frac{k}{m}}(\mathbf{y}) = \frac{1}{k}\overline{\theta}_k(\mathbf{y}), \quad \text{for } k = 1,2,\ldots,m. \tag{16}$$

Note that for $k = 1$, $M_{\frac{1}{m}}(\mathbf{y}) = \overline{\theta}_1(\mathbf{y}) = \theta_1(\mathbf{y}) = M(\mathbf{y})$ thus representing the maximum outcome, and for $k = m$, $M_1(\mathbf{y}) = \frac{1}{m}\overline{\theta}_m(\mathbf{y}) = \frac{1}{m}\sum_{i=1}^m \theta_i(\mathbf{y}) = \frac{1}{m}\sum_{i=1}^m y_i = \mu(\mathbf{y})$ which is the mean outcome. Except for these two limiting cases, the definition (16) is hardly implementable due to the use of the ordering operator. The following theorem shows that the worst conditional mean can be found by minimization of a scalar piecewise linear convex function.

Theorem 1. For any vector $\mathbf{y} \in R^m$ the corresponding quantity $\overline{\theta}_k(\mathbf{y})$ represents the minimum value of the (scalar) optimization:

$$\overline{\theta}_k(\mathbf{y}) = \min_{t\in R} \tfrac{1}{m}\sum_{i=1}^{m}\Big[k(t-y_i)_+ + \\ +(m-k)(y_i-t)_+\Big] + \tfrac{k}{m}\sum_{i=1}^{m} y_i \tag{17}$$

while $\overline{t} = \theta_k(\mathbf{y})$ is an optimal solution (argument) of the above optimization.

Proof. First, we show that $\overline{t} = \theta_k(\mathbf{y})$ minimizes the function:

$$g_k(t) = \sum_{i=1}^{m}\Big[k(t-y_i)_+ + (m-k)(y_i-t)_+\Big]. \tag{18}$$

Note that $g_k(t) = \sum_{i=1}^{m}\Big[k(t-\theta_i(\mathbf{y}))_+ + (m-k)(\theta_i(\mathbf{y})-t)_+\Big]$. Consider $t = \overline{t} + \delta$ with any $\delta \in R$ (positive or negative). For $i = k+1,\ldots,m$

$$\big((\overline{t}+\delta) - \theta_i(\mathbf{y})\big)_+ \geq \big(\overline{t} - \theta_i(\mathbf{y})\big)_+ + \delta$$

and

$$\big(\theta_i(\mathbf{y}) - (\overline{t}+\delta)\big)_+ \geq 0 \ ,$$

while for $i = 1,\ldots,k$

$$(\theta_i(\mathbf{y}) - (\overline{t}+\delta))_+ \geq (\theta_i(\mathbf{y}) - \overline{t})_+ - \delta$$

and

$$\big((\overline{t}+\delta) - \theta_i(\mathbf{y})\big)_+ \geq 0 \ .$$

Hence, one gets

$$k\sum_{i=1}^{m}\big((\overline{t}+\delta) - y_i\big)_+ \geq k\sum_{i=1}^{m}(\overline{t}-y_i)_+ + k(m-k)\delta$$

and

$$(m-k)\sum_{i=1}^{m}\big(y_i - (\overline{t}+\delta)\big)_+ \geq (m-k)\sum_{i=1}^{m}(y_i-\overline{t})_+ + \\ -(m-k)k\delta.$$

Thus finally, $g_k(\overline{t}) \leq g_k(\overline{t}+\delta)$ for all $\delta \in R$.

Further, calculating the minimal value of (18), we get:

$$\begin{aligned} g_k\big(\theta_k(\mathbf{y})\big) &= \\ &= \sum_{i=1}^{m}\Big[k\big(\theta_k(\mathbf{y}) - y_i\big)_+ + (m-k)\big(y_i - \theta_k(\mathbf{y})\big)_+\Big] = \\ &= k\sum_{i=k+1}^{m}\big(\theta_k(\mathbf{y}) - \theta_i(\mathbf{y})\big) - (m-k)\sum_{i=1}^{k}\big(\theta_k(\mathbf{y}) - \theta_i(\mathbf{y})\big) = \\ &= k\sum_{i=1}^{m}\big(\theta_k(\mathbf{y}) - \theta_i(\mathbf{y})\big) - m\sum_{i=1}^{k}\big(\theta_k(\mathbf{y}) - \theta_i(\mathbf{y})\big) = \\ &= km\theta_k(\mathbf{y}) - k\sum_{i=1}^{m}\theta_i(\mathbf{y}) - mk\theta_k(\mathbf{y}) + m\sum_{i=1}^{k}\theta_i(\mathbf{y}) = \\ &= m\sum_{i=1}^{k}\theta_i(\mathbf{y}) - k\sum_{i=1}^{m}\theta_i(\mathbf{y}) = m\overline{\theta}_k(\mathbf{y}) - k\sum_{i=1}^{m} y_i\,. \end{aligned}$$

Hence

$$\overline{\theta}_k(\mathbf{y}) = \frac{1}{m}g_k\big(\theta_k(\mathbf{y})\big) + \frac{k}{m}\sum_{i=1}^{m} y_i = \frac{1}{m}\min_{t\in R} g_k(t) + \frac{k}{m}\sum_{i=1}^{m} y_i$$

which completes the proof of (17).

It follows from Theorem 1 that, for a given vector $\mathbf{y}$, the value of $\overline{\theta}_k(\mathbf{y})$ may be found by solving the linear program:

$$\overline{\theta}_k(\mathbf{y}) = \min \sum_{i=1}^{m}\Big(\frac{k}{m}d_i^- + \frac{m-k}{m}d_i^+\Big) + \frac{k}{m}\sum_{i=1}^{m} y_i$$

$$\text{subject to}$$

$$d_i^+ - d_i^- = y_i - t, \quad d_i^+, d_i^- \geq 0 \ \forall i,$$

where t is an unbounded variable representing a freely selected target while nonnegative variables d_i^+ and d_i^- represent, for several outcome values y_i, their upside and downside deviations from the selected target t, respectively. Moreover, the target variable t takes the value of $\theta_k(\mathbf{y})$ at the optimal solution. The linear program can be further simplified by the elimination of variables d_i^- representing the downside deviations. Hence, following (16), the worst k/m–conditional mean $M_{\frac{k}{m}}(\mathbf{y})$, for $k = 1,2,\ldots,m$, is given by the following optimization:

$$M_{\frac{k}{m}}(\mathbf{y}) = \min\Big\{t + \frac{1}{k}\sum_{i=1}^{m} d_i^+ : d_i^+ \geq y_i - t,\ d_i^+ \geq 0 \forall i\Big\}. \tag{19}$$

This allows us to define *the k/m–conditional minimax* solution for the unweighted allocation problem (12) as the optimal solution to the optimization problem:

$$\min\Big\{t + \frac{1}{k}\sum_{i=1}^{m} d_i^+ : \mathbf{x} \in Q; \quad d_i^+ \geq f_i(\mathbf{x}) - t,\ d_i^+ \geq 0 \ \forall i\Big\} \tag{20}$$

or simply

$$\min\Big\{t + \frac{1}{k}\sum_{i=1}^{m}\big(f_i(\mathbf{x}) - t\big)^+ : \mathbf{x} \in Q\Big\},$$

where $(.)^+$ denotes the nonnegative part of a number.

One may notice that formula (17) in Theorem 1 as well as the subsequent optimization problems (19) or (20) defining the conditional minimax, all they are given directly on outcomes y_i without any use of the ordering operator Θ. Thus, in the case of a weighted allocation problem, Theorem 1 applied to the corresponding disaggregated problem (with equal weights) results in formulas allowing us to reaggregate the outcomes related to the same services. Hence, in the presence of demand weights $w_i > 0$, for any real tolerance level $0 < \beta \leq 1$, there is well defined the *worst* β*–conditional mean*

$$M_\beta(\mathbf{y}) = \min\left\{t + \frac{1}{\beta}\sum_{i=1}^{m} \bar{w}_i d_i^+ : d_i^+ \geq y_i - t,\ d_i^+ \geq 0\ \forall i\right\},$$

where $\bar{w}_i$ denote the normalized weights (1). This allows us to define the β*–conditional minimax* solution for the weighted allocation problem as the optimal solution to the following problem:

$$\min\left\{t + \frac{1}{\beta}\sum_{i=1}^{m} \bar{w}_i d_i^+ : \mathbf{x} \in Q;\ d_i^+ \geq f_i(\mathbf{x}) - t,\ d_i^+ \geq 0\ \forall i\right\}. \tag{21}$$

Note that (21) uses $m+1$ auxiliary variables and m inequalities to define minimization of the worst conditional mean. When the tolerance level β tends to 0, then all the deviational variables d_i^+ are forced to 0. Therefore, the limiting problem of the standard minimax optimization takes the simpler form (14). On the other hand, for $\beta = 1$, problem (21) takes the form

$$\min\left\{\sum_{i=1}^{m} \bar{w}_i(d_i^+ + t) : \mathbf{x} \in Q;\ d_i^+ + t \geq f_i(\mathbf{x}),\ d_i^+ \geq 0\ \forall i\right\},$$

which can be simplified to the standard minisum optimization (13).

The cumulative ordered outcomes (15), used to introduce the worst conditional mean, are closely related with the Pigou-Dalton theory of inequality measurement [12] and the Lorenz curves. Assume that the allocation problem (12) is transformed (disaggregated) into the unweighted one (that means all the demand weights are equal to 1). Vector $\bar{\Theta}(\mathbf{y})$ $\left(\text{exactly } \frac{1}{m}\bar{\Theta}(\mathbf{y})\right)$ can be viewed graphically with the curve connecting point (0,0) and points $(i/m, \bar{\theta}_i(\mathbf{y})/m)$ for $i = 1, 2, \ldots, m$. Graphs of vectors $\bar{\Theta}(\mathbf{y})$ take the form of unnormalized concave curves, the *(upper) absolute Lorenz curves*.

The absolute Lorenz curves defines the relation (partial order) of the equitable dominance. The equitable dominance is originally defined by axioms of efficiency, impartiality and the Pigou-Dalton principle of transfers [7, 10]. Nevertheless, due to the results of the majorization theory [7], it can be expressed with inequalities on the absolute Lorenz curves. Exactly, outcome vector $\mathbf{y}' \in Y$ equitably dominates $\mathbf{y}'' \in Y$, if and only if $\bar{\theta}_i(\mathbf{y}') \leq \bar{\theta}_i(\mathbf{y}'')$ for all $i \in I$ where at least one strict inequality holds. We say that an allocation pattern $\mathbf{x} \in Q$ is *equitably efficient* (is an equitably efficient solution of the multiple criteria problem (12)), if and only if there does not exist any $\mathbf{x}' \in Q$ such that $\mathbf{f}(x')$ equitably dominates $\mathbf{f}(\mathbf{x})$. Note that with the relation of equitable dominance an outcome vector of small unequal outcomes may be preferred to an outcome vector with large equal outcomes. Each equitably efficient solution is also an efficient solution but not *vice verse*.

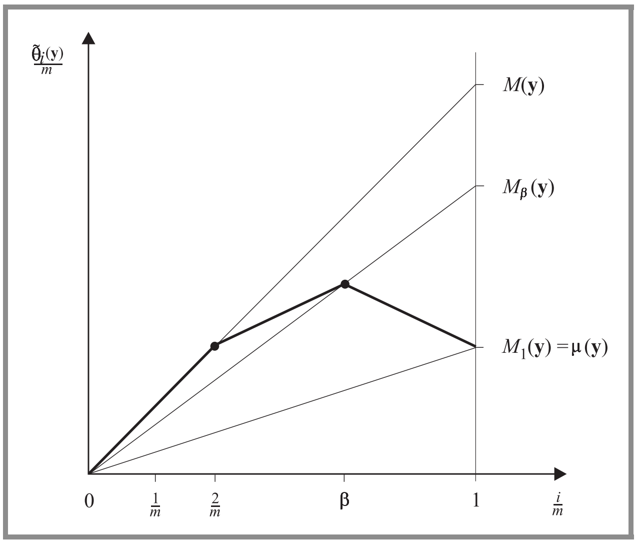


Fig. 1. Absolute Lorenz curve and the worst conditional means.

Recall that the worst conditional mean is defined as $M_{\frac{k}{m}}(\mathbf{y}) = \bar{\theta}_k(\mathbf{y})/k$ while vector $\frac{1}{m}\bar{\Theta}(\mathbf{y})$ can be viewed graphically with the upper absolute Lorenz curve connecting point (0,0) and points $(i/m, \bar{\theta}_i(\mathbf{y})/m)$ for $i = 1, 2, \ldots, m$. Hence, as shown in Fig. 1, the worst conditional mean represents the projection of the point of the Lorenz curve onto the vertical line at point 1 ($i = m$). This also demonstrates that for any given outcome vector $\mathbf{y}$, the worst conditional mean $M_\beta(\mathbf{y})$ is monotonic (nonincreasing), when considered as a function of β, i.e. $0 < \beta' \leq \beta'' \leq 1$ implies $M_{\beta'}(\mathbf{y}) \geq M_{\beta''}(\mathbf{y})$. Further, since the worst conditional mean $M_\beta(\mathbf{y})$ is a quantity proportional to the value of the absolute Lorenz curve at a specific point β, comparison of the worst conditional means (for the same given β) is consistent with the equitable dominance. Exactly, this leads to the following assertion.

Theorem 2. Except for allocation patterns with identical the worst conditional means $M_\beta(\mathbf{y})$, every allocation pattern $\mathbf{x} \in Q$ that is minimal for $M_\beta\left(\mathbf{f}(\mathbf{x})\right)$ is an equitably efficient solution of the allocation problem (12).

4. Computational results

In this section we report some results of our initial computational experience with the conditional minimax solution concept applied to traffic engineering problems. We have solved randomly generated problems following the formulation from Section 2. Thus, our analysis is limited to a simple allocation model where the hubs are arranged in a ring and the traffic engineering problem needs to take into account the bidirectional ring-loading issues.

Our computational tests are based on the randomly generated problems (6)–(11). The generation procedure works as follows. First, a ring with a given number of hubs is built. The clockwise and counterclockwise inter-hub links are distinguished. The delays for these links are generated as random integers uniformly distributed between 5 and 10. Having the ring defined, a given number of services is randomly generated. For each service i, the source node $s(i)$ as well as the destination node $d(i)$ are linked to uniquely selected hub each. The pair of hubs for the given service is chosen randomly from all hubs in the ring, excluding the case of the source node and the destination node attached to the same hub. The delays of links between the source or the destination nodes and their respective hubs in the ring are randomly generated as integers uniformly distributed between 10 and 20. Finally, the demands w_i for the services are generated as random integers uniformly distributed between 1 and 100. All the inter-hub links are assumed to have the same bandwidth. The bandwidth value is defined as a result of the following procedure. We start with initial bandwidth defined as $\sum_{i=1}^{m} w_i$ to guarantee the feasibility (solvability) of the generated problem. Further, we try to reduce the bandwidth still preserving the feasibility. For this purpose, 8 steps of the bisection procedure is applied whereas the current bandwidth is decreased or increased depending on the feasibility of the problem. This allows us to built nontrivial feasible bidirectional ring-loading problems.

The solution concept of conditional minimax provides a compromise between the minimax and the minisum approaches. Table 1 shows the quality of this compromise. It provides average percentage distribution of delays for conditional minimax solutions obtained by varying tolerance level β in the objective M_β. Distribution is calculated as an average of 100 randomly generated problems with 20 hubs and 8 services. Resulting delays are partitioned into clusters of range ten: $[20;30)$, $[30;40)$ etc. Each row represents average distribution for a particular tolerance level β. Exactly, each field gives the percentage of delays within a given range in 100 optimal solutions. It is clear that percentage of low delays increases with β (left columns). On the other hand, for small values of the tolerance level β, the percentage of large delays is forced to zero. With β increasing, large delays begin to occur, first incidentally like delays over 120 for $\beta = 0.1$ or 0.2 (resulting in average percentage below 0.1%), next with a raising percentage.

Table 1
Average distributions of delays for 100 random problems

β	Average percentage of delays for β–conditional minimax solutions												
	20	30	40	50	60	70	80	90	100	110	120	130	140
0.1	2.3	9.0	9.3	8.3	10.5	12.2	12.8	15.9	16.3	3.4	0.0		
0.2	2.1	8.9	8.8	8.4	10.0	10.9	16.3	15.8	15.8	2.9	0.0		
0.3	2.3	8.6	7.9	7.6	11.5	14.5	15.3	16.8	11.7	3.3	0.3	0.1	
0.4	2.3	8.6	8.4	7.9	11.8	15.4	14.3	16.6	10.7	3.5	0.5	0.1	
0.5	2.3	8.6	9.0	8.3	12.8	15.7	13.9	14.2	10.6	4.1	0.5	0.1	
0.6	2.3	8.6	10.7	8.9	12.3	15.4	12.9	13.1	9.9	5.1	0.6	0.2	
0.7	2.3	9.2	11.3	11.3	12.3	14.2	10.3	11.1	11.2	5.2	1.1	0.2	0.2
0.8	2.3	10.1	11.9	12.2	12.9	11.9	8.7	11.3	11.2	5.7	1.3	0.2	0.2
0.9	2.3	10.6	13.4	11.0	11.9	12.2	8.3	10.9	11.9	5.8	1.1	0.4	0.2
1.0	2.3	10.8	13.6	10.8	11.6	12.2	8.3	10.4	12.3	5.9	0.9	0.6	0.2

The main properties of the conditional minimax solution concepts are visible in averages for 100 problems. Nevertheless, a single problem allows us to demonstrate much better the differences among the solutions. Therefore, we have selected and analyzed in details one of the randomly generated problems. Table 2 shows the resulting distributions of delays for four various conditional minimax solution concepts applied to this sample problem. One may notice that the distributions of delays are significantly different despite their means are quite close.

Table 2
Distributions of delays for a sample problem

β	Percentage distribution of delays								μ	M
	30	40	50	60	70	80	100	120		
0.1	22.3	0.2	5.1	22.6	16.7	33.0			65.64	85.35
0.4	22.3	0.2	5.1		50.9	21.4			66.00	85.35
0.7	22.3	0.2	39.3		16.7		21.4		65.43	101.74
1.0	22.3	34.4	5.1		16.7			21.4	64.46	124.86

The distributions of delays generated by several solutions from Table 2 are also presented graphically. In Fig. 2, for each of four distributions of delays the values of all the worst conditional means are shown as functions of the tolerance level. This results in four curves, each starting from the corresponding maximum delay and reaching the mean delay when the tolerance level tends to 1. One may notice that among our four solutions the 0.4-conditional minimax has the smallest worst conditional mean for tolerance levels between 0.21 and 0.6 as well as it remains an alternative optimal solution to the minimax solution for smaller tolerance levels.

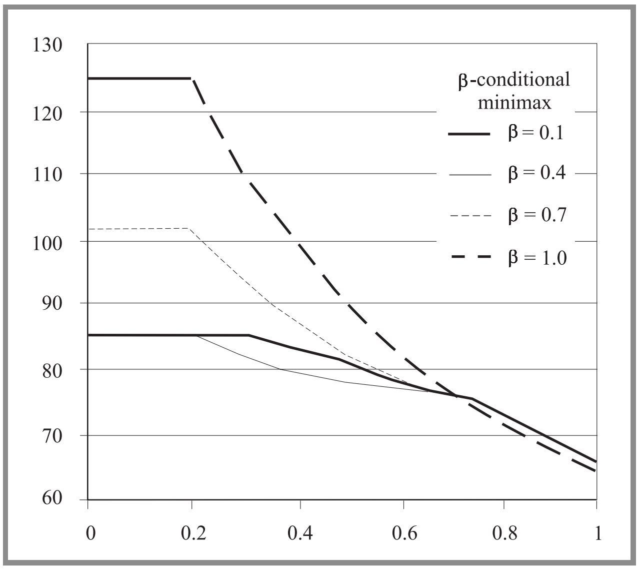


Fig. 2. Curves of the worst conditional means.

Figure 3 shows the absolute Lorenz curves built for the distributions of delays for our four conditional minimax solutions of the sample problem. One might notice from Table 2 that the 0.4–conditional minimax generates the same maximum delay as the 0.1-conditional minimax while its mean is greater than that of the latter. Hence, while dealing with only two criteria of the maximum delay and the mean delay, the 0.4-conditional minimax solution is dominated.

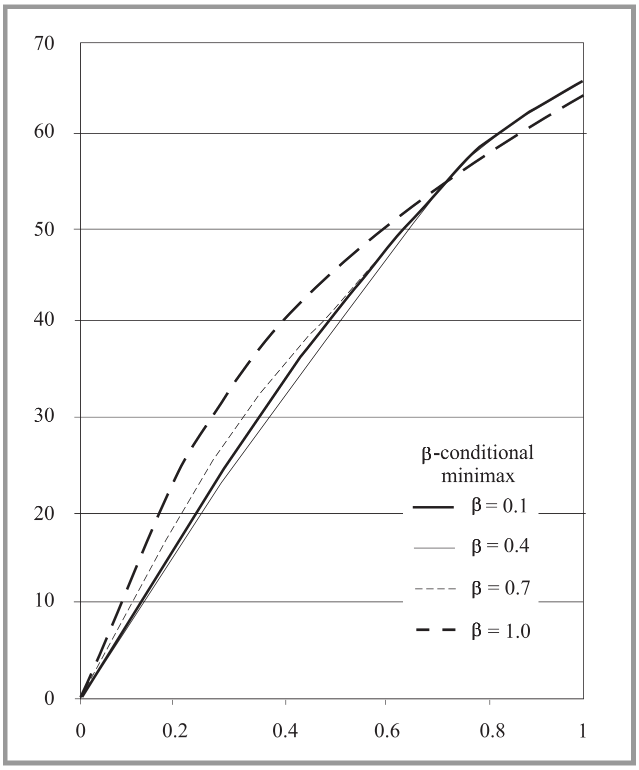


Fig. 3. Absolute Lorenz curves for the sample problem.

Nevertheless, as shown with the absolute Lorenz curves, it is equitably nondominated and the minimization of the worst conditional mean with the tolerance level β between 0.21 and 0.6 points out this solution as optimal.

Table 3
Average solution times for the 0.5-conditional minimax

Hubs	Number of services (m)			
p	50	100	200	500
50	0.10	0.40	0.60	3.20
100	0.40	0.60	1.40	5.40
200	0.40	1.20	2.40	11.00
500	1.20	3.40	6.40	38.60

We tested solution times for different number of services m and number of hubs p. For each specified size parameters we generated randomly 5 problems (6)–(11). The 0.5-conditional minimax solutions were then found. All computations were performed on a PC with the Pentium 200 MHz processor employing the CPLEX 6.0 package [5]. The results are presented in Table 3. Every reported time is an average of 5 results (in seconds) for problems of the given size. One may notice that even problems with 500 hubs were solved very fast.

5. Concluding remarks

Resource allocation problems are concerned with the allocation of limited resources among competing services or other activities so as to achieve the best overall performances. In various systems which serve many users, like in telecommunication systems, there is a need to allocate resources equitably among the competing services. In this paper we have developed an equitable solution concept of the conditional minimax. Although similar to the standard minimax approach, the conditional minimax takes into account the amount of services related to the worst performances. For a specified tolerance level (portion of services amount) β we take into account the entire group of the β portion maximum results and we consider their average as the worst conditional mean to be minimized. According to this definition the solution concept is based on averaging restricted to the group of the worst performances defined by the tolerance level. Hence, by the selection of the tolerance level various equitable preferences may be modeled.

The solution concept of the conditional minimax, similar to the standard minimax approach, can be defined by optimization of a linear objective and a number of auxiliary linear inequalities. Therefore, the concept may be effectively applied to various resource allocation problems. Our initial computational experiments with the conditional minimax applied to a straightforward traffic engineering model (restricted to a single ring bidirectional loading) confirm the theoretical properties of the solution concept. Bidirectional ring loading problems containing up 500 hubs were solved very fast with the general purpose LP solver. Nevertheless, many specific large-scale allocation models (especially discrete ones) may need some specialized exact or approximate algorithms. Thus, further research on computational aspects of the conditional minimax solution concept is necessary.

References

[1] J. F. Campbell, "Hub location and the p-hub median problem", *Oper. Res.*, vol. 44, pp. 923–935, 1996.

[2] D. W. Corne, M. J. Oates, and G. D. Smith, Eds., *Telecommunications Optimization: Heuristics and Adaptive Techniques.* New York: Wiley, 2000.

[3] S. Cosares, D. N. Deutsch, I. Saniee, and O. J. Wasem, "SONET toolkit: a decision support system for designing robust and cost-effective fiber-optic networks", *Interfaces*, vol. 25, pp. 20–40, 1995.

[4] E. Gourdin, "Optimizing Internet networks", *OR/MS Today*, vol. 28, pp. 46–49, 2001.

[5] ILOG Inc., Using the CPLEX Callable Library. Incline Village: ILOG Inc., CPLEX Division, 1997.

[6] J. G. Klincewicz, "Hub location in backbone/tributary network design: a review", *Loc. Sci.*, vol. 6, pp. 307–335, 1998.

[7] M. M. Kostreva and W. Ogryczak, "Linear optimization with multiple equitable criteria", *RAIRO Oper. Res.*, vol. 33, pp. 275–297, 1999.

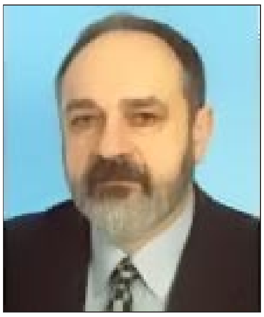
[8] H. Luss, "On equitable resource allocation problems: a lexicographic minimax approach", *Oper. Res.*, vol. 47, pp. 361–378, 1999.

[9] W. Ogryczak, "On the lexicographic minimax approach to location problems", *Eur. J. Oper. Res.*, vol. 100, pp. 566–585, 1997.

[10] W. Ogryczak, "Inequality measures and equitable approaches to location problems", *Eur. J. Oper. Res.*, vol. 122, pp. 374–391, 2000.

[11] J. Rawls, *The Theory of Justice*. Cambridge: Harvard University Press, 1971.

[12] A. Sen, *On Economic Inequality*. Oxford: Clarendon Press, 1973.



Włodzimierz Ogryczak is a professor of Optimization and Decision Support in the Institute of Control and Computation Engineering (ICCE) at the Warsaw University of Technology, Poland. He received both his M.Sc. (1973) and Ph.D. (1983) in mathematics from Warsaw University, and D.Sc. (1997 in computer science from Polish Academy of Sciences. His research interests are focused on theoretical research, computer solutions and interdisciplinary applications in the area of optimization and decision making with the main stress on: multiple criteria optimization and decision support, decision making under risk, location and distribution problems. He has published three books and numerous research articles in international journals. Since 2000 with ICCE, arlier with Warsaw University (Institute of Informatics).

e-mail: W.Ogryczak@ia.pw.edu.pl
Institute of Control & Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Tomasz Śliwiński
e-mail: T.Sliwinski@elka.pw.edu.pl
Institute of Control & Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland