Journal of Applied Mathematics

Fair Optimization and Networks: Models, Algorithms, and Applications

Guest Editors: Wlodzimierz Ogryczak, Hanan Luss, Dritan Nace, and Michał Pióro

Fair Optimization and Networks: Models, Algorithms, and Applications

Journal of Applied Mathematics

Fair Optimization and Networks: Models, Algorithms, and Applications

Guest Editors: Wlodzimierz Ogryczak, Hanan Luss, Dritan Nace, and Michał Pióro

Editorial Board

Contents

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 340913, 3 pages
http://dx.doi.org/10.1155/2014/340913

Editorial

Fair Optimization and Networks: Models, Algorithms, and Applications

Wlodzimierz Ogryczak,[1] Hanan Luss,[2] Dritan Nace,[3] and Michał Pióro[4,5]

[1] *Institute of Control and Computation Engineering, Warsaw University of Technology, 00-665 Warsaw, Poland*
[2] *Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA*
[3] *Laboratoire Heudiasyc, Université de Technologie de Compiègne, 60203 Compiègne, France*
[4] *Department of Electrical and Information Technology, Lund University, 22100 Lund, Sweden*
[5] *Institute of Telecommunications, Warsaw University of Technology, 00-665 Warsaw, Poland*

Correspondence should be addressed to Wlodzimierz Ogryczak; w.ogryczak@elka.pw.edu.pl

Received 21 August 2014; Accepted 21 August 2014; Published 10 September 2014

Optimization models related to the design and evaluation of system policies are mainly focused on efficiency metrics such as the response time, queue length, throughput, and cost. However, in systems which serve many users there is a need to respect some fairness rules while looking for the overall efficiency [1]. Essentially, fairness is an abstract sociopolitical concept that implies impartiality, justice, and equity. In order to ensure fairness in a system, all system entities have to be adequately provided with the system's services. Nevertheless, fair treatment of all entities does not imply equal allocation of services due to constraints imposed on the system by various entities and by the environment. Within the system analysis, fairness was usually quantified with the so-called inequality measures such as variance and mean absolute difference [2]. Unfortunately, direct minimization of typical inequality measures contradicts the maximization of individual outcomes and it also may lead to inefficient designs [3]. Yet, fair optimization with a preference structure that complies with both efficiency and the equitability can be used to generate a variety of fair and efficient solutions [4, 5]. The so-called lexicographic maximin (or minimax) optimization concept [6–9] (and a closely related max-min fairness optimization concept [10]) extends max-min optimization models and is widely applied to various systems. A lexicographic maximin objective optimizes the worst performance among all system entities, followed by optimizing the second worst performance without degrading the worst one, and so forth. However, this may cause a dramatic worsening of the overall system efficiency. Therefore, several other fair optimization models, which compromise between fairness and overall system efficiency, have been extensively analyzed.

The issue of fairness is widely recognized in location and allocation analysis of public services, where the clients of a system are entitled to fair treatment according to community regulations. The need of fair optimization arises also in more general problems of resource allocation [11], for example, in problems related to various networking systems. Fair network optimization issues are in focus of diverse applications and optimization problems in communication networks [12, 13]. These issues are closely related to situations where it is desirable to achieve an equitable allocation for given resources shared by competing traffic demands. Fairness, and more specifically, lexicographic maximin optimization [6, 8, 14], max-min fairness [10], and proportional fairness [15] are widely studied in the communication network literature, especially in relation to bandwidth allocation and rate adaptation and congestion control in TCP (transmission control protocol) networks. At the same time, these concepts are widely applicable in different settings in network optimization and more specifically in the multicommodity flow network related applications.

The presented special issue strives to serve as a platform to present advances in the field. We have invited potential authors to submit original research articles that propose new models, algorithms, and applications of fair network optimization. Out of a total number of 36 papers that were submitted, after a strict review process, 11 high-quality papers are published. The special issue contains one review paper and ten research papers that consider several closely related and interesting topics. In what follows we briefly describe the published contributions.

The review paper "*Fair optimization and networks: a survey*" authored by the editors jointly with A. Tomaszewski overviews fair optimization models and algorithms supporting efficient and fair resource allocation applicable to problems involving network flows that express realizations of competing activities. The presentation applies to communication systems, power distribution systems, transportation systems, logistics systems, and so forth. A particular focus is on allocation problems related to communication networks since in this area the fair optimization concepts have been extensively developed and widely applied.

The paper "*Optimization of power allocation for a multibeam satellite communication system with interbeam interference*" by H. Wang et al. considers power allocation among multiple beams in satellite communication that compromises between maximizing the total system capacity and providing a fair power allocation among the beams. The model is formulated as a nonlinear optimization problem that considers interbeam interferences. Locally optimal solutions are obtained by employing an iterative procedure that is based on duality theory. Simulation results demonstrate the effectiveness of the proposed allocation algorithm.

The paper "*Constructing fair destination-oriented directed acyclic graphs for multipath routing*" by K. Kalinowska-Górska and F. S. Donado examines the issue of determining candidate paths between node pairs that can potentially be used in multicommodity network flow algorithms with applications in communication networks. The paper determines a subset of candidate directed paths from many sources to a single root node such that the number of paths assigned to each source-root pair is considered fair. This is achieved by formulating and solving an optimization problem with integer variables and a lexicographic maximin objective. The authors provide numerical results that compare their method to several existing methods.

In the paper "*Max-min fair link quality in WSN based on SINR*" A. Gogu et al. address the problem of scheduling max-min fair link transmissions in wireless sensor networks, jointly with transmission power assignment. Given a set of concurrently transmitting links, the considered optimization problem seeks for transmission power levels at the nodes so that the signal-to-interference and noise ratio (SINR) values of active links satisfy the max-min fairness property. By guaranteeing a fair transmission medium (in terms of SINR), other network requirements, such as the scheduling length, the throughput (directly dependent on the number of concurrent links in a time slot), and energy savings (no collisions and retransmissions), can be directly controlled.

In his paper "*Design of optical wireless networks with fair traffic flows*," A. Tomaszewski studies a network design problem under variable/uncertain capacity. The work is motivated by practical considerations in wireless networks such as difficulties in traffic transmissions caused by bad weather conditions. In addition to the nominal network state where capacities have their nominal values, the author introduces a set of failure states where capacities are fixed to lower values. Then, the demands are routed in the network with reduced link capacities by decreasing the packet rate fairly, provided that it does not go below some given threshold.

The paper "*Max-min fairness in WMNs with interference cancelation using overheard transmissions*" written by M. Żotkiewicz deals with a recent topic related to wireless mesh networks, namely, to the interference cancellation (IC) technique. The idea is to use IC even for cases with high SINR values taking advantage of overheard traffic; the proposed idea is illustrated with an example. A relevant MIP model for determining scheduling and maximal throughput in a network under these hypotheses is proposed. Simulation results illustrate the benefits of the proposed approach.

The paper "*Threshold accepting heuristic for fair flow optimization in wireless mesh networks*," by J. Hurkała and T. Śliwiński, focuses once again on wireless mesh networks and reports an application of a list-based threshold accepting (LBTA) heuristic that maximizes total throughput while preserving fairness among network flows. The authors use weighted ordered weighted averaging (WOWA) operator to model fairness, while LBTA heuristic is used instead of simulated annealing (SA). Numerical results show that LBTA performs much better than SA in terms of the computing time.

In "*Fair optimization of video streaming quality of experience in LTE networks using distributed antenna systems and radio resource management*," E. Yaacoub and Z. Dawy focus on QoE (quality of experience) used to measure the multimedia experience of mobile users. The authors propose QoE metrics in order to capture the overall performance of the radio resource management (RRM) algorithms in terms of video quality perceived by the end users. Instead of investigating QoE on the link level, they study metrics for assessing the QoE performance over the whole network, taking into account fairness constraints. They consider the use of distributed antenna systems (DAS) to enhance the performance and show by numerical tests that combining DAS and fair RRM algorithms can lead to significant and fair QoE enhancements for all the users in the network.

The paper "*An approximation algorithm for the facility location problem with lexicographic minimax objective*" by L. Buzna et al. presents a new approximation algorithm to the lexicographic minimax optimum of the discrete facility location problem. The approach uses algorithms originally designed to solve the p-median problem and it allows for finding equitable location of facilities serving a large number of customers.

In the paper "*A fairness relation based on the asymmetric Choquet integral and its application in network resource allocation problems*," A. Honda and M. Köppen study a recent problem of network resource allocation where pairs of users could be in a favourable situation, given that the allocation scheme is refined by some add-on technology. They propose a computational approach based on the framework of relational optimization. For representing different weightings for different pairs of users, a binary relation using the asymmetric Choquet integral is introduced as the most suitable approach to represent fairness.

The paper "*Price of fairness on networked auctions*" by M. Kaleta examines execution of auctions under certain networking constraints applicable, for example, in the electricity, communications, and water allocation problems. In classical auctions, a marginal pricing principle is usually applied, which provides a solution acknowledged by all participants as fair. In the case of networked auctions the fairness conditions can be disrupted and the uniform market pricing cannot be used. The paper focuses on multicommodity networks with sealed-bid auctions and shows that the minimal price of fairness can be achieved if the unconstrained market price settlements are adjusted with additional node-dependent prices paid at each node.

Acknowledgments

The editors would like to thank all the authors and reviewers for their valuable contribution to this special issue. The Polish editors would also like to kindly acknowledge the role of the National Science Centre (Poland) who sponsored their activities related to preparing this special issue through the Grant 2011/01/B/ST7/02967 "Integer programming models for joint optimization of link capacity assignment, transmission scheduling, and routing in fair multicommodity flow networks."

Wlodzimierz Ogryczak
Hanan Luss
Dritan Nace
Michał Pióro

References

[1] H. Luss, *Equitable Resource Allocation: Models, Algorithms, and Applications*, John Wiley & Sons, Hoboken, NJ, USA, 2012.

[2] A. B. Atkinson, "On the measurement of inequality," *Journal of Economic Theory*, vol. 2, pp. 244–263, 1970.

[3] W. Ogryczak, "Inequality measures and equitable locations," *Annals of Operations Research*, vol. 167, pp. 61–86, 2009.

[4] M. M. Kostreva and W. Ogryczak, "Linear optimization with multiple equitable criteria," *RAIRO Operations Research*, vol. 33, no. 3, pp. 275–297, 1999.

[5] M. M. Kostreva, W. Ogryczak, and A. Wierzbicki, "Equitable aggregations and multiple criteria analysis," *European Journal of Operational Research*, vol. 158, no. 2, pp. 362–377, 2004.

[6] H. Luss, "On equitable resource allocation problems: a lexicographic minimax approach," *Operations Research*, vol. 47, no. 3, pp. 361–378, 1999.

[7] H. Luss and D. R. Smith, "Resource allocation among competing activities: a lexicographic minimax approach," *Operations Research Letters*, vol. 5, no. 5, pp. 227–231, 1986.

[8] D. Nace and J. B. Orlin, "Lexicographically minimum and maximum load linear programming problems," *Operations Research*, vol. 55, no. 1, pp. 182–187, 2007.

[9] W. Ogryczak, "On the lexicographic minimax approach to location problems," *European Journal of Operational Research*, vol. 100, no. 3, pp. 566–585, 1997.

[10] D. Nace and M. Pióro, "Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.

[11] W. Ogryczak, "Multicriteria models for fair resource allocation," *Control and Cybernetics*, vol. 36, no. 2, pp. 303–332, 2007.

[12] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.

[13] M. Pióro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, San Francisco, Calif, USA, 2004.

[14] W. Ogryczak, M. Pióro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problem," *Journal of Telecommunications and Information Technology*, vol. 2005, no. 3, pp. 43–56, 2005.

[15] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 612018, 25 pages
http://dx.doi.org/10.1155/2014/612018

Review Article

Fair Optimization and Networks: A Survey

Wlodzimierz Ogryczak,[1] Hanan Luss,[2] Michał Pióro,[3,4] Dritan Nace,[5] and Artur Tomaszewski[4]

[1] *Institute of Control and Computation Engineering, Warsaw University of Technology, 00-665 Warsaw, Poland*
[2] *Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, USA*
[3] *Department of Electrical and Information Technology, Lund University, 22100 Lund, Sweden*
[4] *Institute of Telecommunications, Warsaw University of Technology, 00-665 Warsaw, Poland*
[5] *Laboratoire Heudiasyc, Université de Technologie de Compiègne, 60203 Compiègne, France*

Correspondence should be addressed to Wlodzimierz Ogryczak; w.ogryczak@elka.pw.edu.pl

Received 7 March 2014; Revised 3 August 2014; Accepted 6 August 2014; Published 3 September 2014

Academic Editor: Yuri N. Sotskov

Optimization models related to designing and operating complex systems are mainly focused on some efficiency metrics such as response time, queue length, throughput, and cost. However, in systems which serve many entities there is also a need for respecting fairness: each system entity ought to be provided with an adequate share of the system's services. Still, due to system operations-dependant constraints, fair treatment of the entities does not directly imply that each of them is assigned equal amount of the services. That leads to concepts of fair optimization expressed by the equitable models that represent inequality averse optimization rather than strict inequality minimization; a particular widely applied example of that concept is the so-called lexicographic maximin optimization (max-min fairness). The fair optimization methodology delivers a variety of techniques to generate fair and efficient solutions. This paper reviews fair optimization models and methods applied to systems that are based on some kind of network of connections and dependencies, especially, fair optimization methods for the location problems and for the resource allocation problems in communication networks.

1. Introduction

System design and optimization often lead to diverse allocation problems where limited means must be assigned to competing agents or activities so as to achieve the best overall system performance. Depending on the context, the allocation decisions may pertain to costs, tasks, goods, or other resources that can be assigned to one or several agents (actually most allocation problems can be interpreted as resource allocation problems). Such problems arise in numerous applications of considerable complexity with system components being users, stakeholders and their coalition systems, economic and governmental institutions, policy systems, environmental systems [1, 2], and so forth. Very often complex systems that involve resource allocation can essentially be treated as systems of systems [3, 4].

The generic resource allocation problem may be stated as follows. Each activity is measured by an individual performance function that depends on the resource levels assigned to that activity. A larger function value is considered better, like in the case when the performance is measured in terms of assigned system capacity, quality of service level, service amount available, and so forth. In practical applications, one can distinguish different variants of the general allocation problem depending on whether the resource is divisible or not. In particular, one-to-one allocation of indivisible resources lead to the well-known assignment problem, while many-to-many allocation problems arise in task scheduling where a task can be assigned in parallel to several agents with each agent being potentially in charge of several tasks.

Most approaches to allocation problems are focused on efficiency-based objectives. However, the maximization of either total or average results across all relevant agents may require compromising individual agents for the good of others, as long as everyone's good is taken impartially into account. Thus, with the increasing awareness of system inequity resulting from solely pursuing efficiency, a number of fairness or equity oriented approaches have been developed: a particular example is models of resource allocation that try to achieve some form of fairness in resource allocation patterns [5]. In general, the models relate to the optimization of systems which serve many users and the quality of service provided to every individual user defines the optimization criteria. That pattern applies among others to telecommunication and Internet networks: in those networks it is important to allocate network resources, such as available bandwidth, so as to provide equitable performance to all services and all origin-destination pairs of nodes [6, 7]. Still, there are many other pressing examples of systems where fair distribution of resources is required. Problems of efficient and fair resource allocation arise in complex systems of systems when the system combines a number of component systems such as resource supply systems, utilization systems at demand sites or users, stakeholders and their coalition systems, economic and governmental institutions, policy systems, and environmental systems. Actually, addressing fairness in particular types of systems of systems has become a great challenge of the 21 century [8] as fairly dividing limited natural resources (such as the fossil fuels, the clean water, and the environments capacity to absorb greenhouse gases) is perceived as being of utmost importance.

Essentially, fairness is an abstract sociopolitical concept that implies impartiality, justice and equity. In order to ensure fairness in a given system, all system entities have to be equally well provided with the system's services. For example, the issue of equity is widely recognized in the analysis of locating public services, where the clients of a system are entitled to fair treatment according to community regulations. In that context, the decisions often concern the placement of service centers or other facilities at such positions that all users are treated in an equitable way with respect to certain criteria [9]. In particular, location of the facilities pertaining to public services, such as police and fire departments, and emergency medical facilities, should provide fair response time to all demand locations within a metropolitan area. Similarly, water resources should be allocated fairly [10].

As far as technical systems are concerned, the importance of fairness was early recognized with respect to problems of allocation of bandwidth in telecommunication networks [11, 12] (resulting in many models and methods of fair optimization [7]), flight scheduling [13], and allocation of takeoff and landing "slots" at airports [14]. In such areas as allocation of resources in high-tech manufacturing and optimal allocation of water and energy resources, the context of fair resource allocation was additionally enriched by considering possible substitutions among the resources; models with such substitutions are presented in [5, Ch. 4] and [15–17].

In general, complex systems require mathematical programming models in order to describe the dependencies and to enable system optimization. Many such models are based on some kind of network of connections and dependencies. In particular, wide range of system models are related to some kind of network flows that express realizations of competing activities [18]. This applies to telecommunication systems, power distribution systems, transportation systems, logistics systems, and so forth. The discrete location problems can also be viewed in terms of such network system [19, 20].

The general purpose of this paper is to review fair optimization models and algorithms supporting efficient and fair resource allocation in problems related to such network models. The particular focus is on location-allocation problems and allocation problems related to communication networks since in those areas the fair optimization concepts have been extensively developed and widely applied.

The paper is organized as follows. In the next section we present methodological foundations of fair optimization models. In Section 3, the most important models and methods of fair optimization in communication networks are reviewed. Section 4 aims at reviewing applications of fairness optimization in location and allocation problems. The computational complexity issues are addressed in Section 5. The paper is concluded by addressing the most important directions of the development of fair optimization methodology for network systems.

2. Fairness, Equity, and Fair Optimization

2.1. Efficiency and Equity. The generic allocation problem deals with a system comprising a set I of m services (activities, agents) and a given set Q of allocation patterns (allocation decisions). For each service $i \in I$, a function $f_i(\mathbf{x})$ of the allocation pattern $\mathbf{x} \in Q$ is defined. This function measures the outcome (effect) $y_i = f_i(\mathbf{x})$ of allocation pattern $\mathbf{x}$ for service i. In applications we consider this measure that usually expresses the service quality. In general, outcomes can be measured (modeled) as service time, service costs, and service delays as well as in a more subjective way. In typical formulations a larger value of the outcome 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 maximized which allows us to view the generic resource allocation problem as a vector maximization model. Consider

$$\max \{\mathbf{f}(\mathbf{x}) : \mathbf{x} \in Q\}, \tag{1}$$

where $\mathbf{f}(\mathbf{x})$ is a vector-function that maps the decision space $X = R^n$ into the criterion space $Y = R^m$ and $Q \subset X$ denotes the feasible set. We consider complex systems represented by mathematical programming models and specifically models based on some network of connections and dependencies.

An outcome vector $\mathbf{y}$ is attainable if it expresses outcomes of a feasible solution $\mathbf{x} \in Q$ (i.e., $\mathbf{y} = \mathbf{f}(\mathbf{x})$). The set of all the attainable outcome vectors is denoted by A. Note that, in general, convexity of the feasible set Q and concavity of the outcome function $\mathbf{f}$ do not guarantee convexity of the corresponding attainable set A. Nevertheless, the multiple criteria maximization model (1) can be rewritten in the equivalent form

$$\max \{\mathbf{y} : y_i \le f_i(\mathbf{x})\ \forall i, \mathbf{x} \in Q\}, \tag{2}$$

where the attainable set A is convex whenever Q is convex and functions f_i are concave.

Model (1) only specifies that we are interested in maximization of all objective functions f_i for $i \in I = \{1, 2, \ldots, m\}$.

In order to make it operational, one needs to assume some solution concept specifying what it means to maximize multiple objective functions. The solution concepts may be defined by properties of the corresponding preference model [21]. 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 which may be expressed in terms of the vector inequality.

Simple solution concepts for multiple criteria problems are defined by aggregation (or utility) functions $g : Y \rightarrow R$ to be maximized. Thus, the multiple criteria problem (1) is replaced with the maximization problem. Consider

$$\max \{g(\mathbf{f}(\mathbf{x})) : \mathbf{x} \in Q\}. \tag{3}$$

In order to guarantee the consistency of the aggregated problem (3) with the maximization of all individual objective functions in the original multiple criteria problem (or Pareto-optimality of the solution), the aggregation function must be strictly increasing with respect to every coordinate.

The simplest aggregation functions commonly used for the multiple criteria problem (1) are defined as the total outcome $T(\mathbf{y}) = \sum_{i=1}^{m} y_i$, equivalently as the mean (average) outcome $\mu(\mathbf{y}) = T(\mathbf{y})/m = (1/m)\sum_{i=1}^{m} y_i$ or alternatively as the worst outcome $M(\mathbf{y}) = \min_{i=1,\ldots,m} y_i$. The mean (total) outcome maximization is primarily concerned with the overall system efficiency. As based on averaging, it often provides a solution where some services are discriminated in terms of performance. On the other hand, the worst outcome maximization, that is, the so-called max-min solution concept,

$$\max \left\{ \min_{i=1,\ldots,m} f_i(\mathbf{x}) : \mathbf{x} \in Q \right\}, \tag{4}$$

is regarded as maintaining equity. Indeed, in the case of a simplified resource allocation problem with knapsack constraints, the max-min solution,

$$\max \left\{ \min_{i=1,\ldots,m} y_i : \sum_{i=1}^{m} a_i y_i \le b \right\}, \tag{5}$$

takes the form $\overline{y}_i = b / \sum_{i=1}^{m} a_i$ for all $i \in I$, thus, meeting the perfect equity requirement $\overline{y}_1 = \overline{y}_2 = \cdots = \overline{y}_m$. In the general case, with possible more complex feasible set structure, this property is not fulfilled [22, 23]. Nevertheless, if there exists a Pareto-optimal vector $\overline{\mathbf{y}} \in \mathbf{f}(Q)$ satisfying the perfect equity requirement $\overline{y}_1 = \overline{y}_2 = \cdots = \overline{y}_m$, then $\overline{\mathbf{y}}$ is the unique optimal solution of the max-min problem (4) [24].

Actually, the distribution of outcomes may make the max-min criterion partially passive when one specific outcome is relatively very small for all the solutions. For instance, while allocating clients to service facilities, such a situation may be caused by existence of an isolated client located at a considerable distance from all the facilities. Maximization of the worst service performances is then reduced to maximization of the service performances for that single isolated client leaving other allocation decisions unoptimized. For instance, having four outcome vectors (1, 1, 1), (8, 1, 1), (1, 8, 1), and (8, 8, 1) available, they are all optimal in the corresponding max-min optimization, as the third outcome cannot be better than 1. Maximization of the first and the second outcome is then not supported the max-min solution concept, allowing one to select (1, 1, 1) as the optimal solution. This is a clear case of inefficient solution where one may still improve other outcomes while maintaining fairness by leaving at its best possible value the worst outcome. The max-min solution may be then regularized according to the Rawlsian principle of justice. Rawls [25, 26] considers the problem of ranking different "social states" which are different ways in which a society might be organized taking into account the welfare of each individual in each society, measured on a single numerical scale. Applying the Rawlsian approach, any two states should be ranked according to the accessibility levels of the least well-off individuals in those states; if the comparison yields a tie, the accessibility levels of the next-least well-off individuals should be considered, and so on. Formalization of this concept leads us to the lexicographic maximin optimization model or the so-called max-min fairness where the largest feasible performance function value for activities with the smallest (i.e., worst) performance function value (this is the maximin solution) are followed by the largest feasible performance function value for activities with the second smallest (i.e., second worst) performance function value, without decreasing the smallest value, and so forth. The lexicographic maximin solution is known in the game theory as the nucleolus of a matrix game. It originates from an idea, presented by Dresher [27], to select from the optimal (max-min) strategy set of a player a subset of optimal strategies which exploit mistakes of the opponent optimally. It has been later refined to the formal nucleolus definition [28] and generalized to an arbitrary number of objective functions [29]. The concept was early considered in the Tschebyscheff approximation [30] as a refinement taking into account the second largest deviation, the third one and further to be hierarchically minimized. Actually, the so-called strict approximation problem on compact ordered sets is resolved by introducing sequential optimization of the norms on subspaces. Luss and Smith [31] published the first paper on lexicographic maximin approach for resource allocation problems with continuous variables and multiple resource constraints. Within the communications or network applications the lexicographic maximin approach has appeared already in [11, 12] and now under the name max-min fair (MMF) is treated as one of the standard fairness concepts [7]. The lexicographic maximin has been used for general linear programming multiple criteria problems [32–34], as well as for specialized problems related to multiperiod resource allocation with and without substitutions [5, Ch. 5] and [35–39].

In discrete optimization it has been considered for various problems [40, 41] including the location-allocation ones [42]. Luss [43] presented an expository paper on equitable resource allocations using a lexicographic minimax (or lexicographic maximin) approach while [44] provides wide discussion of various models and solution algorithms in connection with communication networks. The recent book by Luss [5] brings together much of the equitable resource allocation research from the past thirty years and provides current state of art in models and algorithm within wide gamut of applications.

Actually, the original introduction of the MMF in networking characterized the MMF optimal solution by the lack of a possibility to increase of any outcome without decreasing of some smaller outcome [12]. In the case of convex attainable set (as considered in [12]) such a characterization represents also lexicographic maximin solution. In nonconvex case as pointed out in [45] such strictly defined MMF solution may not exist while the lexicographic maximin always exists and it covers the former if it exists (see [46] for wider discussion). Therefore, the MMF is commonly identified with the lexicographic maximin while the classical MMF definition is considered rather as an algorithmic approach which is applicable only for convex models. We follow this in the remainder of the paper. Indeed, while for convex problems it is relatively easy to form sequential algorithms to execute lexicographic maximin by recursive max-min optimization with fixed smallest outcomes (see [5, 31–33, 43, 44, 46, 47]), for nonconvex problems the sequential algorithms must be built with the use of some artificial criteria (see [24, 40, 42, 44, 48] and [5, Ch. 7]). Some more discussion is provided in Section 2.4.

2.2. From Equity to Fair Optimization. The concept of fairness has been studied in various areas beginning from political economics problems of fair allocation of consumption bundles [25, 49–52] to abstract mathematical formulation [53, 54]. Fairness is, essentially, an abstract sociopolitical concept of distributive justice that implies impartiality and equity in distribution of goods. In order to ensure fairness in a system, all system entities have to be equally well provided with the system's services. Therefore, in systems analysis and operational research fairness was usually quantified with the so-called inequality measures to be minimized [55–60] or fairness indices [61, 62]. Typical inequality measures are some deviation type dispersion characteristics. They are inequality relevant which means that they are equal to 0 in the case of perfectly equal outcomes while taking positive values for unequal ones. The simplest inequality measures are based on the absolute measurement of the spread of outcomes or deviations from the mean, like the mean absolute difference, maximum absolute difference, standard deviation (variance), mean absolute deviation, and so forth. Relative inequality measures are frequently used. For instance, measures are normalizezd by mean outcome like the Gini coefficient, which is the relative mean difference.

Complex systems require usually mathematical programming models in order to describe the dependencies and to make possible system optimization. Many such models are based on some network of connections and dependencies. A wide range of systems models is related to some flows within a network expressing realizations of competing activities [18]. This applies to communication systems, power distribution systems, transportation systems, logistics systems, and so forth. Among others the discrete location problems can be viewed in terms of such network system [19, 20]. Typically fairness is considered in relation to division of a given amount (the cake division problem) imposing a consistency requirement, the reference points must sum to the total amount available to the agents. A methodology capable to model and solve fair allocation problems in the context of system optimization must take into account possible increase of the amount. Unfortunately, direct minimization of typical inequality measures contradicts the maximization of individual outcomes and it may lead to inferior decisions. The max-min fairness represented by lexicographic maximin optimization meets such needs. This specific concept may be generalized to concepts of fairness expressed by the equitable optimization [9, 24, 43, 63–65] representing inequality averse optimization rather than inequality minimization. Since the term equitable optimization or equitable resource allocation is frequently used as limited to the lexicographic maximin optimization (see [5]), we use the term fair optimization to express wider class of equitable approaches.

The concept of fair optimization is a specific refinement of the Pareto-optimality taking into account the inequality minimization according to the Pigou-Dalton approach. First of all, the fairness requires impartiality of evaluation, thus, focusing on the distribution of outcome values while ignoring their ordering. That means that, in the multiple criteria problem (1), 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:

$$\left(y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(m)}\right) \cong \left(y_1, y_2, \ldots, y_m\right) \quad \text{for any permutation } \pi \text{ of } I, \tag{6}$$

which means that any permuted outcome vector is indifferent in terms of the preference relation. 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:

$$y_{i'} > y_{i''} \Longrightarrow \mathbf{y} - \varepsilon \mathbf{e}_{i'} + \varepsilon \mathbf{e}_{i''} \succ \mathbf{y} \quad \text{for} \quad 0 < \varepsilon < y_{i'} - y_{i''}. \tag{7}$$

The rational preference relations satisfying additionally axioms (6) and (7) are called hereafter *fair (equitable) rational preference relations*. We say that outcome vector $\mathbf{y}'$ *fairly (equitably) dominates* $\mathbf{y}''$, if and only if $\mathbf{y}'$ is preferred to $\mathbf{y}''$ for all fair rational preference relations. In other words, $\mathbf{y}'$ fairly dominates $\mathbf{y}''$, if there exists a finite sequence of vectors $\mathbf{y}^j$ $(j = 1, 2, \ldots, s)$ such that $\mathbf{y}^1 = \mathbf{y}''$, $\mathbf{y}^s = \mathbf{y}'$ and $\mathbf{y}^j$ is constructed from $\mathbf{y}^{j-1}$ by application of either permutation of coordinates, equitable transfer, or increase of a coordinate. An allocation pattern $\mathbf{x} \in Q$ is called *fairly (equitably) efficient* or simply *fair* if $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is fairly nondominated. Note that each fairly efficient solution is also Pareto-optimal but not vice verse.

In order to guarantee fairness of the solution concept (3), additional requirements on aggregation (utility) functions need to be introduced. The aggregation function must be symmetric, that is, for any permutation π of I, $g(y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(m)}) = g(y_1, y_2, \ldots, y_m)$ as well as being equitable (to satisfy the principle of transfers) $g(y_1, \ldots, y_{i'} - \varepsilon, \ldots, y_{i''} + \varepsilon, \ldots, y_m) > g(y_1, y_2, \ldots, y_m)$ for any $0 < \varepsilon < y_{i'} - y_{i''}$. Such functions were referred to as (strictly) Schur-concave [66]. In the case of a strictly increasing and strictly Schur-concave function, every optimal solution to the aggregated optimization problem (3) defines some fairly efficient solution of allocation problem (1) [64].

Both simplest aggregation functions, the mean and the minimum, are symmetric although they do not satisfy strictly the equitability requirement. For any strictly concave and strictly increasing utility function $u : R \rightarrow R$, the aggregation function $g(\mathbf{y}) = \sum_{i=1}^{m} u(y_i)$ is a strictly monotonic and equitable, thus, defining a family of the fair aggregations [64]. Consider

$$\max \left\{ \sum_{i=1}^{m} u(f_i(\mathbf{x})) : \mathbf{x} \in Q \right\}. \tag{8}$$

Various concave utility functions u can be used to define the fair aggregations (8) and the resulting fair solution concepts. In the case of the outcomes restricted to positive values, one may use logarithmic function, thus, resulting in the proportional fairness (PF) solution concept [67, 68]. Actually, it corresponds to the so-called Nash criterion [69] which maximizes the product of additional utilities compared to the status quo. Again, in the case of a simplified resource allocation problem with knapsack constraints, the PF solution,

$$\max \left\{ \sum_{i=1}^{m} \log(y_i) : \sum_{i=1}^{m} a_i y_i \le b \right\}, \tag{9}$$

takes the form $\overline{y}_i = b/a_i$ for all $i \in I$, thus, allocating the resource inversely proportional to the consumption of particular activities.

For positive outcomes a parametric class of utility functions,

$$u(y_i, \alpha) = \begin{cases} \dfrac{y_i^{1-\alpha}}{(1-\alpha)}, & \text{if } \alpha \neq 1, \\ \log(y_i), & \text{if } \alpha = 1, \end{cases} \tag{10}$$

may be used to generate various fair solution concepts for $\alpha > 0$ [70]. The corresponding solution concept (8), called α-fairness, represents the PF approach for $\alpha = 1$, while with α tending to the infinity it converges to the MMF. For large enough α one gets generally an approximation to the MMF while for discrete problems large enough α guarantee the exact MMF solution. Such a way to identify the MMF solution was considered in location problems [40, 42] as well as to content distribution networking problems [71, 72]. 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. For a common case of upper bounded outcomes $y_i \le u^*$ one may maximize power functions $-\sum_{i=1}^{m}(u^* - y_i)^p$ for $1 < p < \infty$ which is equivalent to minimization of the corresponding p-norm distances from the common upper bound u^* [64].

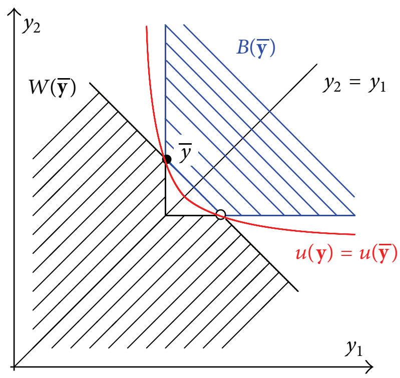


Figure 1: The fair dominance structures $W(\overline{\mathbf{y}})$: the set of outcomes fairly dominated by $\overline{\mathbf{y}}$ and $B(\overline{\mathbf{y}})$: the set of outcomes fairly dominating $\overline{\mathbf{y}}$.

Figure 1 shows the structure of fair dominance for two-dimensional outcome space. For any outcome vector $\overline{\mathbf{y}}$, the fair dominance relation distinguishes set $W(\overline{\mathbf{y}})$ of dominated outcomes (obviously worse for all fair rational preferences) and set $B(\overline{\mathbf{y}})$ of dominating outcomes (obviously better for all fair rational preferences). Some outcome vectors remain neither dominated nor dominating (in white areas) and they can be differently classified by various specific fair solution concepts. The lexicographic maximin assigns the entire interior of the inner white triangle to the set of preferred outcomes while classifying the interior of the external open triangles as worse outcomes. Isolines of various utility functions split the white areas in different ways. For instance, there is no fair dominance between vectors $(1, 100)$ and $(2, 2)$ and the MMF considers the latter as better while the proportional fairness points out the former. On the other hand, vector $(2, 99)$ fairly dominates $(1, 100)$ and all fairness models (including MMF and PF) prefer the former. One may notice that the set $W(\overline{\mathbf{y}})$ of directions leading to outcome vectors being dominated by a given $\overline{\mathbf{y}}$ is, in general, not a cone and it is not convex. Although, when we consider the set $B(\overline{\mathbf{y}})$ of directions leading to outcome vectors dominating given $\overline{\mathbf{y}}$ we get a convex set.

Certainly, any fair solution concept usually leads to some deterioration of the system efficiency when comparing to the sole efficiency optimization. This is referred to as the price of fairness and it was quantified as the relative difference with respect to a fully efficient solution that maximizes the sum of all performance functions (total outcome) [73], that is, the price of fairness concept F on the attainable set A is defined as

$$\mathrm{POF}(F, A) = \frac{\left(\sum_{i=1}^{m} y_i^T - \sum_{i=1}^{m} y_i^F\right)}{\sum_{i=1}^{m} y_i^T}, \tag{11}$$

where $\mathbf{y}^T$ is the outcome vector maximizing the total outcome $T(\mathbf{y})$ on A while $\mathbf{y}^F$ denotes the outcome vector maximizing the fair optimization concept $F(\mathbf{y})$ on A. Formula (11) is applicable only to the problems with a positive total outcome—this, however, is a common case for attainable sets of models based on some network of connections and dependencies. Bertsimas et al. [73] examined the price of fairness for a broad family of problems, focusing on PF and MMF models. They shown that for any compact and convex attainable sets A with equal maximum achievable outcome, which are greater than 0, the price of proportional fairness is bounded by

$$\mathrm{POF}\,(\mathrm{PF}, A) \le 1 - \frac{2\sqrt{m}}{m}, \tag{12}$$

and the price of max-min fairness is bounded by

$$\mathrm{POF}\,(\mathrm{MMF}, A) \le 1 - \frac{4m}{(m+1)^2}. \tag{13}$$

Moreover, the bound under PF is tight if $\sqrt{m}$ is integer, and the bound under MMF is tight for all m. Similar analysis for the α-fairness [74] shows that the price of α-fairness is bounded by

$$\begin{aligned} \mathrm{POF}\,(\alpha F, A) &\le 1 - \min_{\eta\in[1,m]} \frac{\eta^{1+1/\alpha} + m - \eta}{\eta^{1+1/\alpha} + (m-\eta)\,\eta} \\ &\cong 1 - O\left(m^{-\alpha/(\alpha+1)}\right). \end{aligned} \tag{14}$$

The price of fairness strongly depends on the attainable set structure. One can easily construct problems where any fair solution is also maximal with respect to the total outcome (no price of fairness occurs). In [75], the α-fairness concept for network flow problems was analyzed and a class of networks was generated with the property that a fairer allocation is always more efficient. In particular, it implies that max-min fairness may achieve higher total throughput than proportional fairness.

2.3. Multicriteria Models. The relation of fair dominance can be expressed as a vector inequality on the cumulative ordered outcomes [63]. The latter can be 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 $\theta_1(\mathbf{y}) \le \theta_2(\mathbf{y}) \le \cdots \le \theta_m(\mathbf{y})$ and there exists a permutation π of set I such that $\theta_i(\mathbf{y}) = y_{\pi(i)}$ for $i = 1, \ldots, m$. Next, we apply cumulation to the ordered outcome vectors to get the following quantities:

$$\overline{\theta}_i(\mathbf{y}) = \sum_{j=1}^{i} \theta_j(\mathbf{y}) \quad \text{for } i = 1, \ldots, m \tag{15}$$

expressing, respectively, the worst outcome, the total of the two worst outcomes, and the total of the three worst outcomes. Pointwise comparison of the cumulative ordered outcomes $\overline{\Theta}(\mathbf{y})$ for vectors with equal means was extensively analyzed within the theory of equity [76] or the mathematical theory of majorization [66], where it is called the relation of Lorenz dominance or weak majorization, respectively. It includes the classical results allowing to express an improvement in terms of the Lorenz dominance as a finite sequence of Pigou-Dalton equitable transfers. It can be generalized to vectors with various means, which allows one to justify the following statement [63, 77]. Outcome vector $\mathbf{y}' \in Y$ fairly dominates $\mathbf{y}'' \in Y$, if and only if $\overline{\theta}_i(\mathbf{y}') \ge \overline{\theta}_i(\mathbf{y}'')$ for all $i \in I$ where at least one strict inequality holds.

Fair solutions to problem (1) can be expressed as Pareto-optimal solutions for the multiple criteria problem with objectives $\overline{\Theta}(\mathbf{f}(\mathbf{x}))$. Consider

$$\max \left\{ \left(\overline{\theta}_1(\mathbf{f}(\mathbf{x})), \overline{\theta}_2(\mathbf{f}(\mathbf{x})), \ldots, \overline{\theta}_m(\mathbf{f}(\mathbf{x})) \right) : \mathbf{x} \in Q \right\}. \tag{16}$$

Hence, the multiple criteria problem (16) may serve as a source of fair solution concepts. Note that the aggregation maximizing the mean outcome corresponds to maximization of the last objective $\overline{\theta}_m(\mathbf{f}(\mathbf{x}))$ in problem (16). Similarly, the max-min corresponds to maximization of the first objective $\overline{\theta}_1(\mathbf{f}(\mathbf{x}))$. As limited to a single criterion they do not guarantee the fairness of the optimal solution. On the other hand, when applying the lexicographic optimization to problem (16),

$$\operatorname{lex}\max \left\{ \left(\overline{\theta}_1(\mathbf{f}(\mathbf{x})), \overline{\theta}_2(\mathbf{f}(\mathbf{x})), \ldots, \overline{\theta}_m(\mathbf{f}(\mathbf{x})) \right) : \mathbf{x} \in Q \right\}, \tag{17}$$

one gets the lexicographic maximin solution concept, that is, the classical equitable optimization model [5] representing the MMF.

For modeling various fair preferences one may use some combinations of the criteria in problem (16). In particular, for the weighted sum aggregation on gets $\sum_{i=1}^{m} s_i \overline{\theta}_i(\mathbf{y})$, which can be expressed with weights $\omega_i = \sum_{j=i}^{m} s_j$ $(i = 1, \ldots, m)$ allocated to coordinates of the ordered outcome vector, that is, as the so-called ordered weighted average (OWA) [78, 79]:

$$\max \left\{ \sum_{i=1}^{m} \omega_i \theta_i(\mathbf{f}(\mathbf{x})) \ : \mathbf{x} \in Q \right\}. \tag{18}$$

If weights ω_i are strictly decreasing and positive, that is, $\omega_1 > \omega_2 > \cdots > \omega_{m-1} > \omega_m > 0$, then each optimal solution of the OWA problem (18) is a fairly efficient solution of (1). Such OWA aggregations are sometimes called ordered ordered weighted averages (OOWA) [80]. When looking at the structure of fair dominance (Figure 1), the piece-wise linear isolines of the OOWA split the white areas of outcome vectors remaining neither dominated nor dominating (cf. Figure 2).

When differences between weights tend to infinity, the OWA model becomes the lexicographic maximin [81]. On the other hand, with the differences between subsequent monotonic weights approaching 0, the OWA model tends to the mean outcome maximization while still preserving fair optimizations properties (cf. Figure 3).

To the best of our knowledge, the price of fairness related to the fair OWA models has not been studied till now. The OWA aggregation may model various preferences from the max to the min. Yager [78] introduced a well appealing concept of the andness measure to characterize the OWA

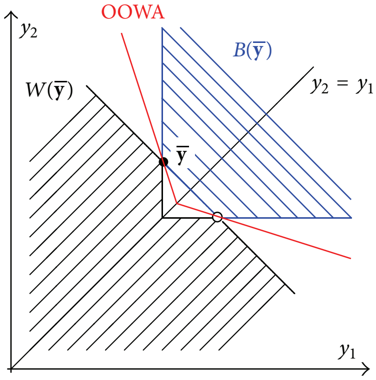


Figure 2: The fair dominance structure and the ordered OWA optimization.

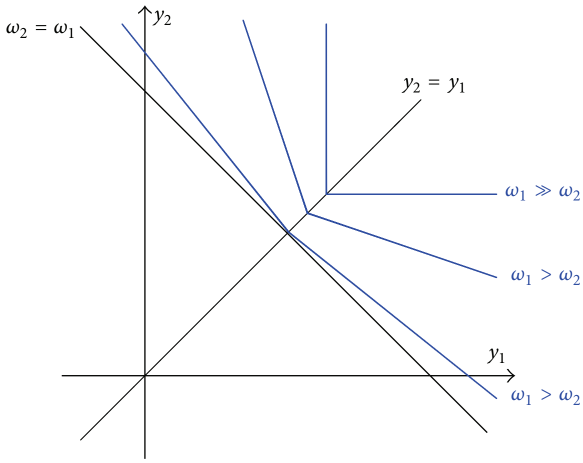


Figure 3: Variety of fair OWA aggregations.

operators. The degree of andness associated with the OWA operator is defined as

$$\text{andness}(\omega) = \frac{\sum_{i=1}^{m}((m-i)/(m-1))\,\omega_i}{\sum_{i=1}^{m}\omega_i}. \tag{19}$$

For the min aggregation representing the OWA operator with weights $\omega = (1, 0, \ldots, 0)$ one gets andness$(\omega) = 1$ while for the max aggregation representing the OWA operator with weights $\omega = (0, \ldots, 0, 1)$ one has andness$(\omega) = 0$. For the total (mean) outcome one gets andness$((1/m, 1/m, \ldots, 1/m)) = 1/2$. OWA aggregations with andness greater than 1/2 are considered fair, and fairer when andness gets closer to 1. A given andness level does not define a unique set of weights ω. Various monotonic sets of weights with a given andness measure may be generated (cf., [82, 83] and references therein).

The definition of quantities $\overline{\theta}_k(\mathbf{y})$ is complicated as requiring ordering. Nevertheless, the quantities themselves can be modeled with simple auxiliary variables and linear constraints. Although, maximization of the kth smallest outcome is a hard (combinatorial) problem. The maximization of the sum of k smallest outcomes is a linear programming (LP) problem as $\overline{\theta}_k(\mathbf{y}) = \max_t(kt - \sum_{i=1}^{m}\max\{t - y_i, 0\})$ where t is an unrestricted variable [84, 85]. This allows one to implement the OWA optimization quite effectively as an extension of the original constraints and criteria with simple linear inequalities [86] (without binary variables used in the classical OWA optimization models [87]) as well as to define sequential methods for lexicographic maximin optimization of discrete and nonconvex models [48]. Various fairly efficient solutions of (1) may be generated as Pareto-optimal solutions to multicriteria problem:

$$\max \quad (\eta_1, \eta_2, \ldots, \eta_m) \tag{20a}$$

$$\text{s.t.} \quad \mathbf{x} \in Q, \tag{20b}$$

$$\eta_k = kt_k - \sum_{i=1}^{m} d_{ik}, \quad k = 1, \ldots, m, \tag{20c}$$

$$t_k - d_{ik} \le f_i(\mathbf{x}), \quad d_{ik} \ge 0, \quad i, k = 1, \ldots, m. \tag{20d}$$

Recently, the duality relation between the generalized Lorenz function and the second order cumulative distribution function has been shown [88]. The latter can also be presented as mean shortfalls (mean below-target deviations) to outcome targets τ:

$$\delta_\tau(\mathbf{y}) = \frac{1}{m}\sum_{i=1}^{m}(\tau - y_i)_+. \tag{21}$$

It follows from the duality theory [88] that one may completely characterize the fair dominance by the pointwise comparison of the mean shortfalls for all possible targets. Outcome vector $\mathbf{y}'$ fairly dominates $\mathbf{y}''$, if and only if $\delta_\tau(\mathbf{y}') \le \delta_\tau(\mathbf{y}'')$ for all $\tau \in R$ where at least one strict inequality holds. In other words, the fair dominance is equivalent to the increasing concave order more commonly known as the Second Stochastic Dominance (SSD) relation [89].

For m-dimensional outcome vectors we consider, all the shortfall values are completely defined by the shortfalls for at most m different targets representing values of several outcomes y_i while the remaining shortfall values follow from the linear interpolation. Nevertheless, these target values are dependent on specific outcome vectors and one cannot define any universal grid of targets allowing to compare all possible outcome vectors. In order to take advantages of the multiple criteria methodology one needs to focus on a finite set of target values. Let $\tau_1 < \tau_2 < \cdots < \tau_r$ denote the all attainable outcomes. Fair solutions to problem (1) can be expressed as Pareto-optimal solutions for the multiple criteria problem with objectives $\delta_{\tau_j}(\mathbf{f}(\mathbf{x}))$. Consider

$$\min\left\{\left(\delta_{\tau_1}(\mathbf{f}(\mathbf{x})), \delta_{\tau_2}(\mathbf{f}(\mathbf{x})), \ldots, \delta_{\tau_r}(\mathbf{f}(\mathbf{x}))\right) : \mathbf{x} \in Q\right\}. \tag{22}$$

Hence, the multiple criteria problem (22) may serve as a source of fair solution concepts. When applying the lexicographic minimization to problem (22) one gets the lexicographic maximin solution concept, that is, the classical equitable optimization model [5] representing the MMF. However, for the lexicographic maximin solution concept one simply performs lexicographic minimization of functions counting outcomes not exceeding several targets [42, 48].

Certainly in many practical resource allocation problems one cannot consider target values covering all attainable outcomes. Reducing the number of criteria we restrict opportunities to generate all possible fair allocations. Nevertheless, one may still generate reasonable compromise solutions [24]. In order to get a computational procedure one needs either to aggregate mean shortages for infinite number of targets or to focus analysis on arbitrarily preselected finite grid of targets. The former turns out to lead us to the mean utility optimization models (8). Indeed, classical results of majorization theory [66] relate the mean utility comparison to the comparison of the weighted mean shortages. Actually, the maximization of a concave and increasing utility function u is equivalent to minimization of the weighted aggregation with positive weights $w(\xi) = -u''(\xi)$ (due to concavity of u the second derivative is negative).

2.4. Methodologies for Solving Lexicographic Maximin Problems. Consider the following resource allocation problem:

$$\operatorname{lex}\max_{\mathbf{x}} \ \overline{\Theta}(\mathbf{f}(\mathbf{x})) = \left(f_{i_1}\left(x_{i_1}\right), f_{i_2}\left(x_{i_2}\right), \ldots, f_{i_m}\left(x_{i_m}\right)\right) \tag{23a}$$

$$\text{s.t.} \quad f_{i_1}\left(x_{i_1}\right) \leq f_{i_2}\left(x_{i_2}\right) \leq \cdots \leq f_{i_m}\left(x_{i_m}\right), \tag{23b}$$

$$\sum_{i \in I} a_{ij} x_i \leq b_j, \quad \forall j \in J, \tag{23c}$$

$$l_i \leq x_i \leq u_i, \quad \forall i \in I, \tag{23d}$$

where the performance functions are strictly increasing and continuous, and $a_{ij} \geq 0$, for all i and j. The lexicographic maximization objective function, jointly with the ordering constraints, defines the lexicographic maximin objective function (this is equivalent to defining the objective function using the ordering mapping Θ). Consider Figure 4 which presents a network that serves point-to-point demands between nodes 1 and 2, nodes 3 and 4, and nodes 3 and 5. The numbers on the links are the link capacities, for example, 4 Gbs on links (1, 3). Suppose demand between a node-pair can be routed only on a single path, where this path is given as part of the input; for example, the path selected between nodes 1 and 2 uses links (1, 3) and (3, 2). The problem of finding the lexicographic maximin solution of demand throughputs between various node-pairs subject to link capacity constraints (which serve as the resource constraints) can be formulated by (23a)–(23d).

It turns out that for various performance functions, such as linear functions and exponential functions, the lexicographic maximin solution of (23a)–(23d) is obtained by simple algebraic manipulations of closed-form expressions and the computational effort is polynomial. This facilitates solving very large problems in negligible computing time. For

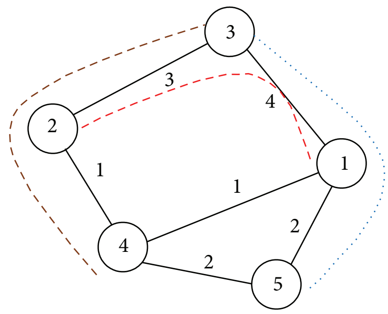


Figure 4: A single path for each demand.

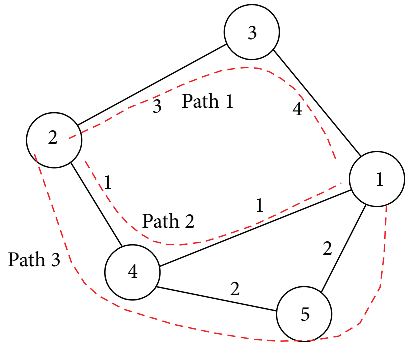


Figure 5: Multiple path for demand between nodes 1 and 2.

other functions, where the solution cannot be derived using closed-form expressions, somewhat more computations are required, in particular, function evaluations complemented by a one-dimensional numerical search are employed (see [5, Ch. 3] and [31, 90, 91]). Algorithms for problem (23a)–(23d) serve as building blocks for more complex problems such as for problems with substitutable resources, for multiperiod problems, and for content distribution problems (see [5, Chs. 4–6]).

Now, consider the cases of performance functions that are nonseparable, where each of the functions $f_i(x_i)$ in (23a) and (23b) is replaced by $f_i(\mathbf{x})$, thus, depending on multiple decision variables. Consider Figure 5 which shows three possible paths for the demand between nodes 1 and 2. The throughput between this node-pair is simply the sum of flows along these three paths.

Even for linear performance functions (e.g., throughputs in communication networks) the computational effort is significantly larger as the algorithm for finding the lexicographic maximin solution requires solving repeatedly linear programming problems (see [5, Chs. 3.4, and 6.2], [7, Ch. 8], and [32, 33, 44, 92]).

Next, consider the case of a nonconvex feasible region, for example, with discrete decision variables. For example, consider a communication network (as in Figure 5) where the demand between any node-pair can flow along multiple paths, but only one of these paths may be selected (here the selected path for each demand is a decision variable). The resulting formulation includes 0-1 decision variables [7]. Again, the objective is to find the lexicographic maximin

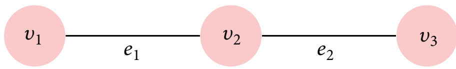


Figure 6: A network example illustrating fairness issues.

solution of the throughputs where each demand uses only one path. All the solution methods above do not apply. If the number of possible distinct outcomes $\tau_1 < \tau_2 < \cdots < \tau_r$ is small, one can construct counting functions, where the kth counting function value is the number of times the kth distinct worst outcome appears in the solution. That means that one introduces functions $\overline{h}_k(\mathbf{y}) = \sum_{l=1}^{k} h_l(\mathbf{y})$ with $h_k(\mathbf{y})$ expressing the number of values τ_k in the outcome vector $\mathbf{y}$. The lexicographic maximin optimization problem is then replaced by lexicographic minimization of the counting functions $\overline{h}_k(\mathbf{y})$ which is solved by repeatedly solving minimization problems with discrete variables:

$$\text{lex min} \quad \left(\sum_{i=1}^{m} z_{1i}, \sum_{i=1}^{m} z_{2i}, \ldots, \sum_{i=1}^{m} z_{r-1,i}\right) \tag{24a}$$

$$\text{subject to} \quad \mathbf{x} \in Q, \tag{24b}$$

$$\tau_{k+1} - f_i(\mathbf{x}) \le M z_{ki}, \quad z_{ki} \in \{0,1\}, \quad i \in I,\ k < r, \tag{24c}$$

where M is a sufficiently large constant (see [5, Ch. 7.2] and [44, 48, 93]). Moreover, in general, binary variables may be eliminated if large numbers of auxiliary continuous variables and constrains are added leading to the formulation based on (22) (see [5, Ch. 7.2] and [44, 48, 93, 94]).

When the number of distinct outcomes is large, we can solve the lexicographic maximin problem by solving lexicographic maximization problems in the format of problems (20a)–(20d) (see [5, Ch. 7.3] and [44, 48, 64, 94–96]). Again, the solution method adds many auxiliary variables and constraints to the formulation.

3. Fairness in Communication Networks

3.1. Fairness and Traffic Efficiency. Fairness issues in communication networks become most profound when dealing with traffic handling. Roughly speaking, whenever the capacity of network resources such as links and nodes is not sufficient to carry the entire offered traffic, a part of the traffic must be rejected. Then a natural question arises: how the total carried traffic traffic should be shared between the network users in a fair way, at the same time assuring acceptable overall traffic carrying efficiency. This kind of problems arise, for example, in the Internet for elastic traffic sources which, from mathematical point of view, can be treated as generating infinite traffic. Thus, the total traffic that can eventually be carried by the network should be fairly split into the traffic flows assigned to individual demands. This issue is illustrated by the following example [7].

Example 1. Consider a simple network composed of two links in series depicted in Figure 6. There are three nodes (v_1, v_2, v_3), two links (e_1, e_2), and three demand pairs ($d_1 = \{v_1, v_2\}$, $d_2 = \{v_2, v_3\}$, $d_3 = \{v_1, v_3\}$). The demands generate elastic traffic, that is, each of them can consume any bandwidth assigned to its path. Suppose that the capacity of the links is the same and equal to 1.5 ($c_1 = c_2 = 1.5$). Let $X = (X_1, X_2, X_3)$ be the path-flows (bandwidth) assigned to demands d_1, d_2, d_3, respectively. Clearly, such a flow assignment is feasible if and only if $X_1, X_2, X_3 \ge 0$ and $X_1 + X_3 \le c_1$, $X_2 + X_3 \le c_2$. For the three basic traffic objectives the solutions are as follows:

(i) max-min fairness (lex max $\Theta(X_1, X_2, X_3)$) : $X_1 = X_2 = X_3 = 0.75$ ($T(X) = 2.25$),

(ii) proportional fairness (max $\log X_1 + \log X_2 + \log X_3$) : $X_1 = X_2 = 1$, $X_3 = 0.5$ ($T(X) = 2.5$), and

(iii) throughput maximization (max $X_1 + X_2 + X_3$) : $X_1 = X_2 = 1.5$, $X_3 = 0$ ($T(X) = 3$).

Above $T(X)$ denotes the throughput, that is, $T(X) = X_1 + X_2 + X_3$. Clearly, the MMF solution is perfectly fair from the demand viewpoint but at the same the worst in terms of throughput. This is because the “long” demand d_3, consuming bandwidth on both links, gets the same flow as the “short” demands d_1, d_2, each consuming bandwidth on its direct link. The PF solution increase the flow of short demands at the expense of the long demand. This is acceptably fair for the demands and increases the throughput. Finally, the $T(X)$ maximization solution is unfair (the long demand gets nothing) but, by assumption, maximizes the throughput.

Note that in this example the price of max-min fairness calculated according to formula (11) is 1/4 which is equal to the upper bound (13). Similarly, the price of proportional fairness 1/6 is close to its upper bound (12). However, the price of fairness strongly depends on the network topology. In [75], the authors demonstrate a class of networks such that an α-fair allocation with higher α is always more efficient in terms of total throughput. In particular, this implies that max-min fairness may achieve higher throughput than proportional fairness.

In the networking literature related to fairness, the above MMF and PF objectives are the most popular. The throughput maximization objective is rarely used, as totally unfair. Instead, a reasonable modification consisting in lexicographical maximization of the two ordered criteria $(\min(X), T(X))$ is used, where $\min(X)$ denotes the minimal element of the demand vector X.

Considering MMF, besides optimization objectives directly related to traffic handling, objectives related to link loads, are commonly considered in communication network optimization. In this case, the traffic volumes of demands to be realized are fixed. We shall come back to this issue later on.

3.2. Generic Optimization Models. The considered network is modeled with a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, undirected or directed, composed of the set of nodes $\mathcal{V}$ and the set of links $\mathcal{E}$. Thus, each link $e \in \mathscr{E}$ represents an unordered pair $\{v, w\}$ (undirected graphs) or an ordered pair (v, w) (directed graphs) of nodes $v, w \in \mathscr{V}$ and is assigned the nonnegative unit capacity cost ξ_e which is a parameter and the maximum capacity $c(e)$ which is a given constant (possibly equal to $+\infty$). When link capacities are subject to optimization, they become optimization variables denoted by $y_e,\ e \in \mathscr{E}$. The cost of the network is given by the quantity $C = \sum_{e\in\mathscr{E}} \xi_e y_e$. The traffic demands are represented by the set $\mathscr{D}$. Each demand $d \in \mathscr{D}$ is characterized by a directed pair $(o(d), t(d))$, composed of the originating node $o(d)$ and the terminating node $t(d)$, and a minimum value $h(d)$ (a parameter, possibly equal to 0) of the traffic volume that has to be carried from $o(d)$ to $t(d)$. Demand volumes and link capacities are expressed in the same units.

Each demand d has a specified set of admissible paths $\mathscr{P}(d)$ (called the path-list) composed of selected elementary paths from $o(d)$ to $t(d)$ in graph $\mathscr{G}$. (Recall that an elementary path does not traverse any node more than once). Paths in $\mathscr{P}(d)$, used to realize the demand (traffic) volumes, are assigned flows $x_p,\ p \in \mathscr{P}(d)$, which are optimization variables. Each value x_p specifies the reference capacity (expressed in the same units as link capacity and demand volume) reserved on path $p \in \mathscr{P}(d)$. The set of all admissible paths is denoted by $\mathscr{P} := \bigcup_{d\in\mathscr{D}} \mathscr{P}(d)$. The maximum path-lists, that is, path-lists $\mathscr{P}(d)$ containing all elementary paths from $o(d)$ to $t(d)$, will be denoted by $\widehat{\mathscr{P}}(d),\ d \in \mathscr{D}$, with $\widehat{\mathscr{P}} := \bigcup_{d\in\mathscr{D}} \widehat{\mathscr{P}}(d)$. The set of all paths in $\mathscr{P}$ traversin a simple network composed of two links in series depicted in Figure 6. There are three nodes (v_1, v_2, v_3), two links (e_1, e_2) and three demand pairs $(d_1 = \{v_1, v_2\}, d_2 = \{v_2, v_3\}, d_3 = \{v_1, v_3\})$. The demands generate elastic traffic, that is, each of them can consume any bandwidth assigned to its path. Suppose that the capacity of the links is the same and equal to 1.5 ($c_1 = c_2 = 1.5$). Let $X = (X_1, X_2, X_3)$ be the path-flows (bandwidth) assigned to demands d_1, d_2, d_3, respectively. g a given link $e \in \mathscr{E}$ will be denoted by $\mathscr{Q}(e)$. Note that in an undirected graph the links can be traversed by paths in both directions while in a directed graph—only in the direction of the link.

Let $X_d = \sum_{p\in\mathscr{P}(d)} x_p$ denote the total flow assigned to demand $d \in \mathscr{D}$, that is, traffic of demand d carried in the network, and let $X = (X_d : d \in \mathscr{D})$. Besides, let $Y_e = \sum_{p\in\mathscr{Q}(e)} x_p$ be the link load induced by the path-flows. Then, the generic feasibility set (optimization space) of a traffic allocation problem (TAP) can be specified as follows:

$$\sum_{p\in\mathscr{P}(d)} x_p = X_d, \quad d \in \mathscr{D}, \tag{25a}$$

$$X_d \geq h(d), \quad d \in \mathscr{D}, \tag{25b}$$

$$\sum_{p\in\mathscr{Q}(e)} x_p = Y_e, \quad e \in \mathscr{E}, \tag{25c}$$

$$Y_e \leq c(e), \quad e \in \mathscr{E}, \tag{25d}$$

$$x_p \in \mathscr{X}, \quad p \in \mathscr{P}. \tag{25e}$$

The set $\mathscr{X}$ specifies the domain of a path-flow variable and is problem-dependent. Two typical cases are $\mathscr{X} = \mathbb{R}_+$ and $\mathscr{X} = \mathbb{Z}_+$. Note that in the undirected graph the path-flows through a link sum up to the link load no matter in which direction they traverse the link.

The three cases of TAP considered in Example 1 above can be now formulated as follows:

(i) TAP/MMF: lex max $\Theta(X)$ subject to (25a)–(25e),

(ii) TAP/PF: max $L(X) = \sum_{d\in\mathscr{D}} \log X_d$ subject to (25a)–(25e), and

(iii) TAP/TM: lex max $(M(X) = \min_{d\in\mathscr{D}} X_d, T(X) = \sum_{d\in\mathscr{D}} X_d)$ subject to (25a)–(25e).

Observe that the third case above is actually different from the third case considered in Example 1 as now throughput maximization is the secondary objective in lexicographical maximization.

When $\mathscr{X} = \mathbb{R}_+$, all the three problems are convex and as such can be approached effectively by means of the algorithms described in [7, 44, 46]. For the TAP/PF version see [67]. In fact, TAP/TM is a two level linear program possibly combined to a single LP [23], and TAP/MMF can be solved as a series of linear programs [32, 33, 44, 97]. Optimization approaches to TAP/PF are presented in [67].

Certainly, the feasible set (25a)–(25e) can be further constrained to consider more restricted routing strategies. The most common restriction is imposed by the single-path requirement that each X_d is carried entirely on one selected path. Then the feasibility set must be augmented by the following constraints:

$$\sum_{p\in\mathscr{P}(d)} u_p = 1, \quad d \in \mathscr{D}, \tag{26a}$$

$$x_p \leq M u_p, \quad p \in \mathscr{P}, \tag{26b}$$

$$u_p \in \{0, 1\}, \quad p \in \mathscr{P}. \tag{26c}$$

In (26a)–(26c), $u_p, p \in \mathscr{P}$ are additional binary routing variables, and M is a "big M" constant. In this setting the above defined TAP problems become essentially mixed-integer programming problems (FTP/PF after a piece-wise approximation of the logarithmic function), and in the case of MMF must be treated by the general approach described in Section 2.3 as problem (20a)–(20d) (see also [44, 48, 64, 94–96] and [5, Ch. 7.3]).

We note that when the routing paths are fixed, that is, when $|\mathscr{P}(d)| = 1,\ d \in \mathscr{D}$, then TAP/MMF becomes the classical fair allocation (equitable resource allocation) problem considered in Section 2.4 (see [12, Sec. 6.5.2] and [5, Ch. 6.1]). This version of the problem can be efficiently solved in polynomial time by the so called water-filling algorithm based on the bottleneck link characterization of the problem (see [45] and Section 3.7). In fact, the bottleneck characterization of this TAP/MMF problem can be directly formulated as an integer programming problem (with binary variables) as demonstrated in [92]. The modular flow version of the problem is considered in [98].

An interesting version of the single-path TAP/MMF problem is considered in [99] that uses the bottleneck formulation of [92]. In that problem, the routes are optimized so to achieve the maximum traffic throughput while maintaining the MMF demand traffic assignment.

The above specified problems use the noncompact link-path formulation where the optimization variables are related to the routing paths. Hence, when we wish to consider all possible elementary paths then the number of variables x_p, $p \in \widehat{\mathscr{P}}$ becomes exponential with the size of the network. In this case path generation algorithm should be applied (this is easy in the case of linear programs) or the problems should be reformulated in the node-link notation using link-flow variables instead of the path-flow variables used in (25a)–(25e).

3.3. Selected Specific Models. In this section we will discuss several specific network optimization models related to various aspects of fairness. An interesting case arise when the traffic demands $h(d)$, $d \in \mathscr{D}$ are considered as given and the design objective is to balance the load of the links, aiming at minimizing the average packet delay in the network. The commonly known formulation of such load balancing is as follows:

$$\min \ r \tag{27a}$$

$$\sum_{p \in \mathscr{P}(d)} x_p = h(d), \quad d \in \mathscr{D}, \tag{27b}$$

$$\sum_{p \in \mathscr{Q}(e)} x_p \leq c(e)\, r, \quad e \in \mathscr{E}, \tag{27c}$$

$$r \in \mathbb{R}, \quad x_p \in \mathscr{X}, \ p \in \mathscr{P}. \tag{27d}$$

Using the MMF notion it is easy to define a load balancing problem, that is stronger than problem (27a)–(27d) which in fact find the maximum element of the MMF vector $R = (r_e : e \in \mathscr{E})$ expressing the relative link loads:

$$\text{lex min} \ \Theta(R) \tag{28a}$$

$$\sum_{p \in \mathscr{P}(d)} x_p = h(d), \quad d \in \mathscr{D}, \tag{28b}$$

$$\sum_{p \in \mathscr{Q}(e)} x_p \leq c(e)\, r_e, \quad e \in \mathscr{E}, \tag{28c}$$

$$\begin{aligned} & r_e \in \mathbb{R}, \quad e \in \mathscr{E}, \\ & x_p \in \mathscr{X}, \quad p \in \mathscr{P}. \end{aligned} \tag{28d}$$

Some variants of the problem given by (28a)–(28d) were studied in [100, 101].

Another version of the MMF load balancing problem (28a)–(28d) maximizes the unused link capacity $\overline{Y} = (\overline{Y}_e : e \in \mathscr{E})$ in a fair way, relevant to circuit switching:

$$\text{lex max} \ \Theta(\overline{Y}) \tag{29a}$$

$$\sum_{p \in \mathscr{P}(d)} x_p = h(d), \quad d \in \mathscr{D}, \tag{29b}$$

$$\sum_{p \in \mathscr{Q}(e)} x_p = Y_e, \quad e \in \mathscr{E}, \tag{29c}$$

$$Y_e \leq c(e), \quad e \in \mathscr{E}, \tag{29d}$$

$$\overline{Y}_e = c(e) - Y_e, \quad e \in \mathscr{E}, \tag{29e}$$

$$x_p \in \mathscr{X}, \quad p \in \mathscr{P}. \tag{29f}$$

Above we have considered flow allocation problems assuming given link capacity. When the link capacity is subject to optimization, that is, when we simultaneously optimize path-flows and link capacities, then we deal with dimensioning problems. An example of such a problem (with a budget constraint) is as follows:

$$\text{lex max} \ \Theta(X) \tag{30a}$$

$$\sum_{p \in \mathscr{P}(d)} x_p = X_d, \quad d \in \mathscr{D}, \tag{30b}$$

$$\sum_{p \in \mathscr{Q}(e)} x_p = Y_e, \quad e \in \mathscr{E}, \tag{30c}$$

$$\sum_{e \in \mathscr{E}} \xi(e)\, Y_e \leq B, \quad e \in \mathscr{E}, \tag{30d}$$

$$x_p \in \mathscr{X}, \quad p \in \mathscr{P}, \tag{30e}$$

where $B > 0$ is a given budget for the total link cost. Note that we have skipped constraint (25b) which has established a lower bound on the demand traffic allocation in formulation (25a)–(25e). If no additional constraints are enforced (as (25b)) then the optimal solution of (30a)–(30e) is trivial. For each demand $d \in \mathscr{D}$, the optimal traffic $X_d = X^*$ is the same and realized on the cheapest path $p(d) \in \widehat{\mathscr{P}}_d$ with respect to the cost $\kappa(d) = \sum_{e \in p(d)} \xi(e)$. Clearly

$$X^* = \frac{B}{\sum_{d \in \mathscr{D}} \kappa(d)}. \tag{31}$$

When the PF objective,

$$\min \sum_{d \in \mathscr{D}} \log X_d, \tag{32}$$

instead of the MMF objective (30a) is considered, then the optimal solution is as follows (see [7, 68, 102]):

$$X_d^* = \frac{B}{\kappa(d)\, |\mathscr{D}|}, \quad d \in \mathscr{D}, \tag{33}$$

so the total optimal flow X_d^* allocated to demand d is inversely proportional to the cost of its shortest path (and allocated to this path).

More complicated optimization problems including link dimensioning were treated in [7, Ch. 13] (see also [103, 104]). For the MMF optimization problems related to wireless networks (in particular, to Wireless Mesh Networks) the reader can refer to [105].

3.4. Extended Fairness Objectives. While the MMF and PF objectives are the most popular in the networking literature related to fairness, there are also attempts to find various fair solutions taking advantages of the multicriteria fair optimization models presented in Section 2.3. In particular, the OWA aggregation (18) was applied to the network dimensioning problem for elastic traffic [95] as well as to the flow optimization in wireless mesh networks [106].

Example 2. Consider the simple network from Example 1 composed of two links in series depicted in Figure 6. There are three demand pairs ($d_1 = \{v_1, v_2\}$, $d_2 = \{v_2, v_3\}$, $d_3 = \{v_1, v_3\}$) generating elastic traffic, where $X = (X_1, X_2, X_3)$ are the path-flows (bandwidth) assigned to demands d_1, d_2, d_3, respectively. Note that the ordered OWA maximization with decreasing weights $\omega = (0.4, 0.35, 0.25)$ results in bandwidth allocation $X_1 = 1.5$, $X_2 = 1.5$, $X_3 = 0$, thus, representing the maximum throughput. Ordered OWA maximization with decreasing weights $\omega = (0.6, 0.3, 0.1)$ results in bandwidth allocation $X_1 = 0.75$, $X_2 = 0.75$, $X_3 = 0.75$ which is the MMF solution.

It was demonstrated that allocations representing the classical fairness concepts (MMF and PF) were easy to achieve [95]. On the other hand, in order to find a larger variety of new compromise solutions it was necessary to incorporate some scaling techniques originating 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 multicriteria problems with discrete (or more general nonconvex) feasible sets, they may fail to identify several compromise efficient solutions. In standard multicriteria optimization, good controllability can be achieved with the direct use of the reference point methodology [107] based on reservation and aspiration levels for each of the activities. The reservation levels are the required activity levels, whereas the aspiration levels are the desired levels, commonly referred to as reference points. The reference point methodology applied to the cumulated ordered outcomes (16) was tested on the problem of network dimensioning with elastic traffic [96, 108]. The tests confirmed the theoretical advantages of the method. Various (compromise) fair solutions for both continuous and modular problems could be easily generated.

Multiple criteria model of the mean shortfalls to all possible targets (22) when applied to network dimensioning problem for elastic traffic results in a model with criteria that measure actual network throughput for various levels (targets) of flows [109]. Thereby, the criteria can easily be introduced into the model. Experiments with the reference point methodology applied to the multiple target throughput model confirmed the theoretical advantages of the method. Various (compromise) fair solutions were easily generated despite the fact that the single path problem (discrete one) was analyzed.

Both the multiple criteria models with the lexicographic optimization of directly defined artificial criteria introduced with some auxiliary variables and linear inequalities provides corresponding implementations for the MMF optimization independently from the problem structure. The approaches guarantee the exact MMF solution for a complete set of criteria and their applicability is limited to rather small networks. In [94] there were developed some simplified sequential approaches with reduced number of criteria, thus, generating effectively approximations to the MMF solutions. Computational analysis on the MMF single-path network dimensioning problems showed the approximated models allowed to solve within a minute problems for networks with 30 nodes and 50 links providing very small approximation errors, thus, suggesting possible usage in many practical applications.

3.5. Fairness on the Session Level. One of the major challenges of the Internet is to provide high performance of data transport. Basically, the problem is how to obtain high utilization of network resources and to ensure required quality of communications services. Those two goals result in a potential trade-off as when the amount of data sent through the network is too high, links become overloaded and the quality of service deteriorates.

The overload occurs when the amount of data loading the outgoing link of the Internet router is higher than the one that can actually be carried. When that happens the link's queue of packets becomes longer, and potentially the queue's buffer finally overflows. That causes the increase of packet delay and delay variations and may also cause packet loss. Both phenomena are perceived by the pair of communicating Internet applications as low quality of data transport.

Let $\mathcal{S}$ be the set of Internet sessions, which are packet flows between pairs of Internet applications. Let function $l : \mathcal{S} \mapsto \mathbb{R}_+$ define the average packet length of the session expressed in bits, and for each $s \in \mathcal{S}$, let variable x_s denote the packet rate of session s. Then, for each $s \in \mathcal{S}$, $x_s l(s)$ is an average bit-rate of session s.

Let $\mathcal{E}$ be the set of network links, and for each $s \in \mathcal{S}$, let $\mathcal{E}(s)$ denote the set of links that are used by session s, and for each $e \in \mathcal{E}$, let $\mathcal{S}(e)$ denote the set of sessions that use link e. Then the load of link $e \in \mathcal{E}$ is equal to $\sum_{s\in\mathcal{S}(e)} x_s l(s)$. Let function $c : \mathcal{E} \mapsto \mathbb{R}_+$ denote the capacity (the bit-rate) of the link. The following constraint expresses the fact that the total load of any link cannot be greater than the link's capacity. Consider

$$\sum_{s\in\mathcal{S}(e)} x_s l(s) \le c(e), \quad e \in \mathcal{E}. \tag{34}$$

The overload of the Internet's link is a very common situation. The links can become overloaded for a number of reasons: when the amount of traffic entering the network becomes significantly larger, when links lose some capacity due to failures, or when they fail completely and the packet flows must be rerouted to some other links that do not have sufficient capacity. Thus, solving the trade-off between utilization and quality of service requires effective mechanisms of handling overload. That is, the place when the concept of fairness is used.

The data between a pair of applications in the Internet can be conveyed using one of two transport protocols, user datagram protocol (UDP) and transport control protocol (TCP). While the UDP is a connectionless data transport protocol, where each data packet is sent individually and there is no interaction between the sending and the receiving application, the TCP protocol is connection-oriented, which means that packets are sent within a connection that must be organized between the sending and the receiving application before the data can be sent, and can be torn down only after the last packet has been delivered. Due to the connection-oriented character of the TCP flows there is an association between the two applications which allows them to control the packet rate.

With the flow control mechanisms of the TCP protocol the rate at which packets are sent is adapted to network conditions: if the amount of available bandwidth is large, packet rate is being increased, and when the links become overloaded the rate is decreased, thus, reducing the overload. The packet rate of the TCP session increases every time the sender application receives an acknowledgement that a packet has reached the destination, and the rate is decreased every time a packet is lost. While the increase is linear, the decrease is geometrical, which helps to ease congestion quickly. In a reactive scenario, the packet is lost when the packet buffer is saturated. In the proactive scenario, to avoid uncontrolled congestion, the random early discard (RED) mechanism of the router can be activated that discards randomly selected packets. However, in both cases a random packet is lost and a randomly selected session is affected.

Arguably, the higher the packet rate of a session the higher the probability that packets of the session will be dropped and the packet rate of the session will be reduced. Thus, if a number of sessions have their packet rate reduced due to congestion of a given link, none of the sessions is supposed to generate packets at an average rate higher than the other sessions. For each $e \in \mathscr{E}$, let variable y_e denote the maximum packet rate on link e. Noticeably, there is some maximum rate at which a particular application can generate packets; let function $r : \mathscr{S} \mapsto \mathbb{R}_+$ define the maximum achievable packet rate of the session. Thus, the packet rate of the session must, potentially, satisfy the following condition:

$$x_s = \min \left\{ r(s), y_e \right\}, \quad e \in \mathscr{E},\ s \in \mathscr{S}(e). \tag{35}$$

Due to (35) the bandwidth of a single link is shared in a fair way. If a link is saturated, every session s attains the same packet rate y_e, unless that rate is higher than the maximum achievable rate $r(s)$ of that session. Thus, the session cannot have packet rate higher than any other session unless the other session's maximum achievable rate is lower than y_e. And only if a link is not saturated, every session attains its maximum achievable packet rate. However, since in general sessions use multiple network links, on a given link a session can in fact have a lower packet rate than other sessions that use that particular link. That results from the fact that the packet rate of the session can be reduced even more due to congestion on some other link. Thus, condition (35) must actually be replaced with the following one:

$$x_s = \min \left\{ r(s), \min_{e \in \mathscr{E}(s)} y_e \right\}, \quad s \in \mathscr{S}. \tag{36}$$

That condition can be interpreted as follows. For any session $s \in \mathscr{S}$ the session's packet rate x_s attempts to approach the maximum achievable packet rate $r(s)$. However, on any link $e \in \mathscr{E}(s)$, that is, used by session s, the value of x_s cannot exceed the maximal packet rate y_e, that is, attained by the sessions that use that particular link. Thus, the session's packet rate x_s can only attain the minimal of those rates $\min_{e \in \mathscr{E}(s)} y_e$ unless that minimal rate is still higher than $r(s)$, in that case the packet rate of s just approaches $r(s)$.

Considering conditions (34) and (36), it can now be seen that the flow control mechanism of the TCP protocol maximizes the vector of the packet rates of individual sessions $\overline{x} \equiv (x_s : \ s \in \mathscr{S})$ in a fair way.Consider

$$\text{lex max} \quad \Theta(\overline{x}), \tag{37a}$$

$$x_s \le r(s), \quad s \in \mathscr{S}, \tag{37b}$$

$$\sum_{s \in \mathscr{S}(e)} x_s l(s) \le c(e), \quad e \in \mathscr{E}, \tag{37c}$$

$$x_s \in \mathbb{R}_+, \quad s \in \mathscr{S}. \tag{37d}$$

The max-min fairness property of the packet rates vector means that the packet rates of the data sessions are increased up to their maximum values unless links become overloaded, and in the case of a link overload, the data sessions on the link decrease their rate to the common highest feasible value. This type of behaviour appears to have far reaching consequences for solving the problem of packet network design that carry elastic traffic when the aim of the design is controlling the quality of services when the capacity of links changes [110].

3.6. Content Distribution Networks. Bandwidth allocation for content distribution through networks composed of multiple tree topologies with directed links and a server at the root of each tree is another problem of fair network optimization [111, 112] and [5, Ch. 6]. Content distribution over networks has become increasingly popular. It may be related, for instance, to a video-on-demand application where multiple programs can be broadcasted from each server. Each server broadcasts along a tree topology, where these trees may share links and each link has a limited bandwidth capacity. Figure 7 presents a network with two trees and servers at the root nodes 1 and 2. The server at node 1 can broadcast programs 1, 2, and 3 and the server at node 2 can broadcast programs 4, 5, and 6. The numbers adjacent to the links are the link capacities and the numbers adjacent to the nodes are the programs requested; for example, links (1, 3) have a capacity of 100 Gbs and programs 2, 3, and 5 are requested at node 7.

These models are fundamentally different from multicommodity network flow models since they do not have flow

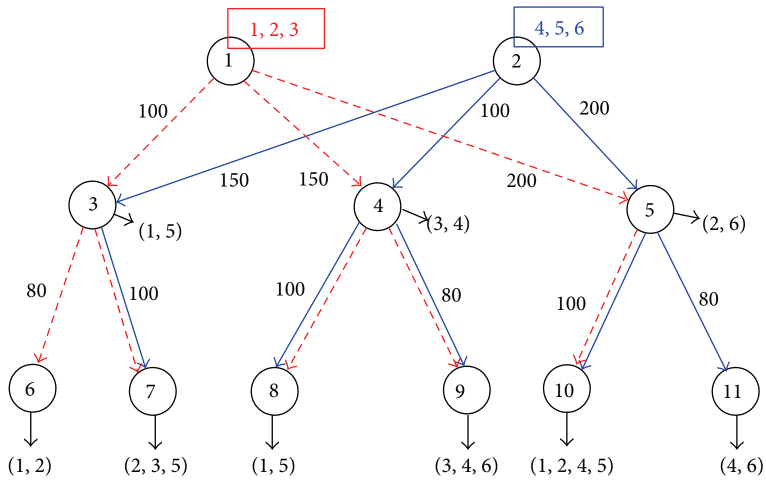


Figure 7: Content distribution from two servers.

conservation constraints as each link carries at most one copy of a program. On the other hand, the models have tree-like ordering constraints for each program as the allocated bandwidth for a given program cannot be increased when moving farther away from the broadcasting server. For each requested program at any node there is associated a performance function that represents satisfaction from the video-on-demand service and depends on the bandwidth available for that program on the incoming link to the node. Fair optimization with respect to all nodes and programs requested performance values is needed. In [111] the MMF model is introduced and a lexicographic max-min algorithm is presented. As shown in [113] the algorithm can be implemented in a distributed mode where most of the computations are done independently and in parallel at all nodes, while some information is exchanged among the nodes. More complex content distribution models and corresponding algorithms are discussed in [114–116].

3.7. Fairness Issues in the IP Traffic. In its beginnings, the Internet suffered from severe deficiencies due to congestion. The answer came from new features added to TCP, namely, employing control admission and additive increase/multiplicative decrease algorithms that led to congestion avoidance and fair rate allocation. The main idea behind was putting the control traffic mechanisms at the end-nodes and combining both packet scheduling with admission control which will lead to fair bandwidth sharing. Plenty of studies have been done on the behavior of the network when such congestion avoidance algorithms are employed. They have shown that this leads to some kind of max-min fair sharing in very simplified networks [117] and to proportional fairness for large networks ([67] etc). This difference is mainly due to end-to-end delays which can be significantly different in large scale networks. At this point, an important topic is how to get close to maximal throughput while keeping a high level of fairness. In [118] there are investigated the performances of networks handling elastic flows (in contrast to stream flows they adjust their rate to the available bandwidth). It is shown that in linear networks under random traffic patterns, ensuring max-min fairness results in better throughput performances comparing to proportional fairness while the converse holds for persistent flows. All these works are situated at the session level and refer to traffic demand as the product of the flow arrival rate with the average flow size. At this stage, a more global solution would come by combining session level decisions (see Section 3.5) with higher level decisions as routing and load balancing. Hence, relations of rate adaptation and congestion control in TCP networks with routing and network design have been the subject of several works over the last decade. Among them, some work has been devoted to the static routing case (connections and corresponding routing paths are given) where source rates are subject to changes. In [12] there is presented the water-filling algorithm for achieving a MMF distribution of resources to connections for the fixed single path routing case (where each connection is associated with a particular fixed path). The main idea behind the algorithm is to uniformly increase the individual allocations of connections until one or more link becomes congested. Then the connections that cannot be improved are removed from the network together with the capacities they occupy; the process continues until all connections are removed. In [119], the problem of MMF bandwidth-sharing among elastic traffic connections when routing is not fixed has been considered in an offline context. The proposed iterative algorithm can be seen as an extension of the water-filling algorithm given in [12] except that the routing is not fixed and at each iteration a new routing is computed while the previously saturated links and the corresponding fair sharing remain fixed until the end of the algorithm.

Load balancing is in a way a problem dual to MMF routing (see TAP/MMF in Section 3.2) as one focuses on min-max fair load sharing instead of max-min fair bandwidth allocation to demands. Achieving load-balancing in a given network consists in distributing the demand traffic (load) fairly among the network links while satisfying a given set of traffic constraints. Fair load sharing means that not only the maximal load among links is minimized, but rather that the sorted (in the nonincreasing order) vector of link loads is minimized lexicographically like in formulation (28a)–(28d). In contrast to the max-min fair routing problem like TAP/MMF, the link load-balancing problem assures fairness in the min-max sense. The problem arises in communication networks when the operator needs to define routing with respect to a given traffic demand matrix such that the network load is fairly distributed among the network links. The problem can be easily modeled and solved by conventional methods in LP using MMF properties for linear link loads. This approach can be applied to more general link load functions (especially nonlinear, frequently used in telecommunications). In practice the load/delay functions considered by network operators are usually nonlinear. A well-known load/delay function, called the Kleinrock function, is given by $f_e/(C_e - f_e)$. It can be shown that any routing achieving min-max fairness for the relative load function (i.e., f_e/C_e) achieves also min-max fair load for the Kleinrock function. This idea is generalized for general link load functions as $(\alpha - 1)^{-1}(1 - f_e/C_e)^{1-\alpha}$ and $(\alpha - 1)^{-1}(f_e - C_e)^{1-\alpha}$, where f_e and C_e give, respectively, the flow and capacity on link e and α is a given constant, see [120] for further details.

The problem of fairness is more complex when dealing with wireless networks and has been addressed in a number of papers during the last decade. A range of problems can be distinguished depending on the network characteristics, from wireless mesh networks, ad-hoc networks, sensor networks, random-access networks, opportunistic ones, and so forth. As for conventional wired networks, a fundamental problem in wireless networks is to estimate their throughput capacity and then to develop protocols to utilize the network close to this capacity without causing congestion in the network and unfairness. Then the first idea that comes in mind to address the fairness problem in wireless networks is the classic approach to manage congestion inherited from wired networks. Then, nodes/flows will have preassigned fair shares and applying admission control would allow ensuring fair sharing. In wireless networks this cannot be applied because of interference, which constrains the set of links that can transmit simultaneously, while in ad-hoc networks nodes and routers mobility renders the problem even more complicated. In WSN (wireless sensor networks) the fairness problem becomes on one hand closely connected to fair data gathering, that is, serving the sources equitably, and on the other hand it is connected to ensuring aware energy consuming because of the reduced energy capacity of nodes in such networks. Then the main constraint that one has to deal with is the so-called MAC (medium access control) constraint. Let us recall briefly what MAC constraint is; we start by its definition. Two basic definitions can be distinguished: the protocol and the physical one. The protocol definition of interference assumes that two links, which are less than k (generally k is taken 2) hops away from each other, interfere potentially and cannot be scheduled in the same time slot. The indicated number of hops refers to the number of hops between the sender nodes of these links. On the other hand, the physical definition is based on the signal-to-interference and noise ratio (SINR) constraint where the transmission links that do not satisfy the SINR constraint cannot be scheduled simultaneously. Hence, this constraint leads to new connected problems namely synchronization and scheduling. Given the above, most of the work related to these strategies is dedicated to scheduling. The basic version of a time slot allocation problem aims to find a slot allocation for all nodes in the network with minimal number of slots such that k hops neighboring nodes are not allocated to the same time slot. The respective optimization problem is the chromatic graph one which aims to minimize the number of colors for coloring the nodes such that two neighbor elements do not use the same color. The problem becomes more difficult if one desires to achieve fairness between connections or sources. All this yields the max-min fair scheduling. In [121] the authors consider scheduling policies for max-min fair allocation of bandwidth in wireless ad-hoc networks. They formalize the max-min fair objective under wireless scheduling constraints and propose a fair scheduling which assigns dynamic weights to the flows such that the weights depend on the congestion in the neighborhood and schedule the flows which constitute a maximum weighted matching. While in [122], the authors propose a quite different alternative. Their method is inspired from per-flow queuing in wired networks and consists of a probabilistic packet scheduling scheme achieving max-min fairness without changing the existing IEEE 802.11 medium access control protocol. When a wireless node is ready to send a packet, the packet scheduler of the node is likely to select the queue whose number of packets sent in a certain time is the smallest and when no packet is available, the transmission is delayed by a fixed duration. In [123] the authors investigate simple queuing models for random traffic and discuss their interest for both wired and wireless transmissions.

With respect to WSN, the rate allocation problem for data aggregation in wireless sensor networks can be posed with two objectives: the first is maximizing the minimum (max-min) lifetime of an aggregation cluster and the second achieving fairness among all data sources. The two objectives cannot be maximized simultaneously and an approach would be to solve recursively first the max-min lifetime for the aggregation cluster, and next the fairness problem. In [124] the authors use this approach and formulate the problem of maximizing fairness among all data sources under a given max-min lifetime, as a convex optimization problem. Next, they compute the optimal rate allocations iteratively by a lexicographic method. In a recent paper [125], the authors address the problem of scheduling MMF link transmissions in wireless sensor networks, jointly with transmission power assignment. Given a set of concurrently transmitting links, the considered optimization problem seeks for transmission power levels at the nodes so that the SINR values of active links satisfy the max-min fairness property. By guaranteeing a fair transmission medium (in terms of SINR), other network requirements, such as the scheduling length, the throughput (directly dependent on the number of concurrent links in a time slot), and the energy savings (no collisions and retransmissions), can be directly controlled.

4. Location and Allocation Problems

4.1. Inequality Measures. The spatial distribution of public goods and services is influenced by facility location decisions and the issue of equity (or fairness) is important in many location decisions. In particular, various public facilities (or public service delivery systems) like schools, libraries, and health-service centers, require some spatial equity while making location-allocation decisions [126, 127]

The generic discrete location problem may be stated as follows. There is given a set of m clients (service recipients). Each client is represented by a specific point. There is also given a set of n potential locations for the facilities and the number (or the maximal number) p of facilities to be located is given ($p \le n$). This means discrete location problems or network location problems with possible locations restricted to some subsets of the network vertices [128]. The main decisions to be made can be described with the binary variables x_j ($j = 1, 2, \ldots, n$) equal to 1 if location j is to be used and equal to 0 otherwise. To meet the problem requirements, the decision variables x_j have to satisfy the following constraints:

$$\sum_{j=1}^{n} x_j = p, \qquad x_j \in \{0, 1\}, \text{ for } j = 1, \ldots, n. \tag{38}$$

where the equation is replaced with the inequality ($\le$) if p specifies the maximal number of facilities to be located. Further the allocation decisions are represented by the additional variables x'_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) equal to 1 if location j is used to service client i and equal to 0 otherwise. The allocation variables have to satisfy the following constraints:

$$\begin{gathered} \sum_{j=1}^{n} x'_{ij} = 1, \quad \text{for } i = 1, \ldots, m. \\ x'_{ij} \le x_j, \quad \text{for } i = 1, \ldots, m, \; j = 1, \ldots, n. \\ x'_{ij} \in \{0, 1\}, \quad \text{for } i = 1, \ldots, m, \; j = 1, \ldots, n. \end{gathered} \tag{39}$$

In the capacitated location problem the capacities of the potential facilities are given which implies some additional constraints.

Let $d_{ij} \ge 0$ ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) denote the distance between client i and location j (travel effort or other effect of allocation client i to location j). For the standard uncapacitated location problem it is assumed that all the potential facilities provide the same type of service and each client is serviced by the nearest located facility. The individual objective functions then can be expressed in the linear form:

$$f_i(\mathbf{x}) = \sum_{j=1}^{n} d_{ij} x'_{ij}, \quad \text{for } i = 1, \ldots, m. \tag{40}$$

These linear functions of the allocation variables are applicable for the uncapacitated as well as for the capacitated facility location problems. In the case of location of desirable facilities a smaller value of the individual objective function means a better effect (smaller distance). This remains valid for location of obnoxious facilities if the distance coefficients are replaced with their complements to some large number: $d'_{ij} = d - d_{ij}$, where $d > d_{ij}$ for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Generally, replacing the distances with their utility values or so-called proximity measures, for example, $u_{ij} = \exp(-\beta d_{ij})$ [129]. Therefore, we can assume that each function f_i is to be minimized as stated in the multiple criteria problem [130].

Further, some additional client weights $w_i > 0$ are included into location model to represent the service demand (or clients importance). Integer weights can be interpreted as numbers of unweighted clients located at exactly the same place. The normalized client weights $\overline{w}_i = w_i / \sum_{i=1}^{m} w_i$ for $i = 1, \ldots, m$ rather than the original quantities w_i. In the case of unweighted problem (all $w_i = 1$), all the normalized weights are given as $\overline{w}_i = 1/m$.

Note that constraints (38) take a very simple form of the binary knapsack problem with all the constraint coefficients equal to 1 [131]. Indeed, the location problem may be viewed as a resource allocation problem on network. It may be considered as capacities allocation to links from an artificial source to potential locations nodes while flows are routed from the source to all client nodes through the the potential location nodes [19, 20].

Equity is usually quantified with the so-called inequality measures to be minimized. Inequality measures were primarily studied in economics [57, 76]. The simplest inequality measures are based on the absolute measurement of the spread of outcomes. Variance is the most commonly used inequality measure of this type and it was also widely analyzed within various location models [132, 133]. However, many various measures have been proposed in the literature to gauge the level of equity in facility location alternatives [58], like the *mean absolute difference* also called the Gini's mean difference [9, 59]. Consider

$$D(\mathbf{y}) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \left| y_i - y_j \right| \overline{w}_i \overline{w}_j, \tag{41}$$

or like the mean absolute deviation

$$\mathrm{MAD}(\mathbf{y}) = \sum_{i=1}^{m} \left| y_i - \mu(\mathbf{y}) \right| \overline{w}_i. \tag{42}$$

In economics one usually considers relative inequality measures normalized by mean outcome. Among many inequality measures perhaps the most commonly accepted is the Gini index (Lorenz measure), $D(\mathbf{y})/\mu(\mathbf{y})$ a relative measure of the mean absolute difference, which has been also analyzed in the location context [134–136]. One can easily notice that a direct minimization of typical inequality measures (especially relative ones) contradicts the minimization of individual outcomes. As noticed by Erkut [134], it is rather a common flaw of all the relative inequality measures that while moving away from the clients to be serviced one gets better values of the measure as the relative distances become closer to one-another. As an extreme, one may consider an unconstrained continuous (single-facility) location problem and find that the facility located at (or near) infinity will provide (almost) perfectly equal service (in fact, rather lack of service) to all the clients. Unfortunately, the same applies to all dispersion type inequality measures, including the upper

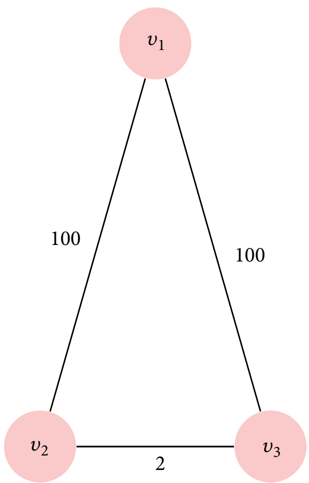


Figure 8: A network location problem for Example 3.

semideviations. This can be illustrated by a simple example of location problem on network.

Example 3. Consider a single facility location on a (triangular) network of 3 nodes: two nodes v_1 and v_2 are close one to the other $d_{12} = 2$, and one remote node v_3 with $d_{13} = d_{23} = 100$ (see Figure 8). Most of the demand is equally distributed in v_1 and v_2. That means the normalized values of weights take values $\overline{w}_1 = \overline{w}_2 = (1-\varepsilon)/2$ and $\overline{w}_3 = \varepsilon$ with a very small positive value ε. While locating facility at node v_1 (or v_2) one gets distance 0 for $(1-\varepsilon)/2$ demand, distance 2 for $(1-\varepsilon)/2$ demand, and large distance 100 for only ε demand. However, $\mu(v_1) = 1 + 99\varepsilon$ and $\text{MAD}(v_1) = (1-\varepsilon)(1+99\varepsilon)$ and in terms of MAD minimization it is beaten by remote location v_3. Indeed, locating facility at node v_3 one gets distance 0 for only ε demand while getting large distance 100 for $(1-\varepsilon)$ demand, thus, much worser than for v_1. Nevertheless, $\mu(v_3) = 100(1-\varepsilon)$ and $\text{MAD}(v_3) = 200\varepsilon(1-\varepsilon)$. Hence, for small values of ε: $\text{MAD}(v_3) < \text{MAD}(v_1)$. Actually, for sufficiently small values ε (e.g., $0 < \varepsilon < 1/200$) location v_3 is a global MAD minimizer on the entire network (when allowing location on edges in addition to the nodes).

For typical inequality measures a simplified bicriteria mean-equity model is computationally very attractive since both the criteria are well defined directly for the weighted location problem without necessity of its disaggregation but it may result in solutions which are inefficient. It turns out that, under the assumption of bounded trade-offs, the bicriteria mean-equity approaches for selected absolute inequality measures (maximum upper deviation, mean semideviation or mean absolute difference) comply with the rules of equitable (fair) optimization [9, 137]. In other words, several inequality measures can be combined with the mean itself into the optimization criteria generalizing the concept of the worst outcome and generating equitably consistent underachievement measures. Simple sufficient conditions for inequality measures to keep this consistency property have been introduced in [137].

This applies, in particular, to the mean absolute difference (41) generating a proper fair solution concept:

$$\begin{aligned} M_{\alpha D}(\mathbf{y}) &= \mu(\mathbf{y}) + \alpha D(\mathbf{y}) \\ &= (1-\alpha)\sum_{i=1}^{m} y_i \overline{w}_i \\ &\quad + \alpha \sum_{i=1}^{m} \sum_{j=1}^{m} \max\{y_i, y_j\}\, \overline{w}_i \overline{w}_j, \end{aligned} \tag{43}$$

for any $0 < \alpha \le 1$. Similar result is valid for the mean absolute deviation (42) but not for the variance [24, 137].

4.2. Lexicographic Minimax and Ordered Medians. Although minimization of the inequality measures contradicts the minimization of individual outcomes, the inequality minimization itself can be consistently incorporated into locational models. The notion of equitable multiple criteria optimization [63] introduces the preference structure that complies with both the outcomes minimization and with the inequality minimization rules [57, 76]. The equitable optimization is well suited for the locational analysis [9, 137, 138]. The equitably (fair) efficiency models presented in Section 2.3 apply also to the minimized outcomes, as commonly considered in location-allocation problems. The equitable minimization can be modeled with the standard multiple criteria optimization applied to the cumulative ordered outcomes, expressing, respectively, the worst outcome, the total of the two worst outcomes, the total of the three worst outcomes, and so forth. However, in the case of minimization the worst outcome means the largest rather than the smallest. Hence, the corresponding model takes form

$$\min\left\{\left(\widehat{\theta}_1(\mathbf{f}(\mathbf{x})), \widehat{\theta}_2(\mathbf{f}(\mathbf{x})), \ldots, \widehat{\theta}_m(\mathbf{f}(\mathbf{x}))\right) : \mathbf{x} \in Q\right\}, \tag{44}$$

where $\widehat{\theta}_i(\mathbf{y}) = \sum_{j=m-i+1}^{m} \theta_j(\mathbf{y})$. The minimax, called the center solution concept, represents only the first criterion, while the total outcome criterion, called the median solution concept, is focused on the last criterion. Several cent-dian solution concepts combining these two criteria have been considered (see [139] and references therein). For unweighted location problems, a compromise solution concept was introduced by Slater [140] as the so-called k-centrum where the sum of the k largest distances is minimized. Consistently with typical distribution characteristics, The k-centrum concept is restricted to unweighted problems. Although some weights are used to scale the specific distances [141] (which may be considered as a definition of distance dependent outcomes), the demand weights as defining the distribution of clients are not considered. Ogryczak and Zawadzki [142] introduced a parametric generalization of the k-centrum concept applied to weighted problems by taking into account the portion of demand related to the largest outcomes (distances) rather than the specific number of worst outcomes. Namely, for a specified portion β of demand the entire β portion (quantile) of the largest outcomes is taken into account and their average is considered as the (worst) conditional β-mean outcome.

According to this definition the concept of conditional median is based on averaging restricted to the portion of the worst outcomes. For the unweighted location problems and $\beta = k/m$, the conditional β-mean represents the average of the k largest outcomes, thus, modeling the k-centrum solution concept.

However, in order to guarantee the equitable efficiency of a selected location pattern one needs to take into account all the ordered outcomes (all the criteria in (44). The entire multiple criteria ordered model is rich with various equitably efficient solution concepts [64, 142, 143]. For the weighted sum aggregation one gets the OWA aggregation $\sum_{i=1}^{m} \omega_i \theta_i(\mathbf{y})$ (18), called the ordered median solution concept [144]. If the OWA weights are strictly increasing and positive, that is $0 < \omega_1 < \omega_2 < \cdots < \omega_{m-1} < \omega_m$, then each optimal solution of the OWA problem (18) is an equitably (fairly) efficient location pattern. Although the cumulated ordered outcomes can be expressed with linear programming models [85], these approaches requires the disaggregation of location problem with the demand weights which usually dramatically increases the problem size.

When applying the lexicographic optimization to problem (44),

$$\text{lex min}\left\{\left(\widehat{\theta}_1(\mathbf{f}(\mathbf{x})), \widehat{\theta}_2(\mathbf{f}(\mathbf{x})), \ldots, \widehat{\theta}_m(\mathbf{f}(\mathbf{x}))\right) : \mathbf{x} \in Q\right\}, \tag{45}$$

one gets the lexicographic minimax solution concept, called also lexicographic center [42] as a lexicographic refinement of the center solution concept. The lexicographic minimax location may be converted to a lexicographic minimization objective by constructing counting functions that count, for each possible distinct outcome, the number of occurrences of the specified outcome. It is quite simple to construct such counting functions for the discrete location problem (see [42, 48] or [5, Ch. 7.2]).

The lexicographic maximin approach can be applied to various location problems. The sensor location problem is an extension of the equitable facility location problem [5, Ch. 7.3] and [145]. Consider a set N of nodes that need to be monitored as protection against undesired intrusion and a set M of nodes where sensors can be placed. Let I_j be the subset of nodes in M that can monitor node $j \in N$, and let K be the number of available sensors that can be placed among nodes in M. The protection level provided to node j is represented by the number of sensors that monitor node j. The sought after solution is the lexicographic maximin solution with respect to the number of sensors that protects nodes $j \in N$. Figure 9 presents a problem with four locations that need to be monitored ($N = \{2, 3, 4, 5\}$) and four locations where sensors can be placed ($M = \{1, 3, 4, 5\}$). The links that connect the nodes represent subsets I_j. Two sensors can be placed among the nodes in M ($K = 2$).

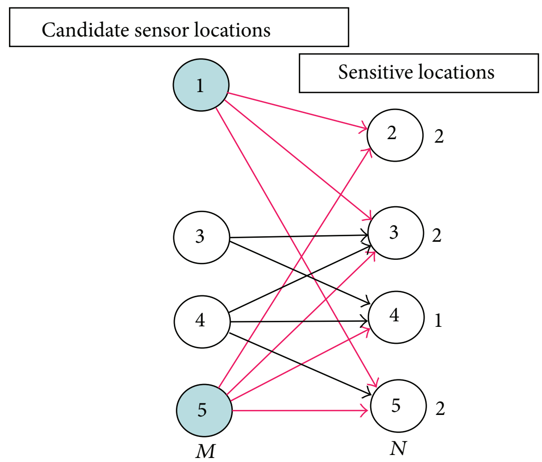


Figure 9: The sensor location problem.

The formulation of this problem is as follows:

$$\text{lex max} \quad \overline{\Theta}(X) = \left(X_{j_1}, X_{j_2}, \ldots, X_{j_n}\right) \tag{46a}$$

$$\text{s.t.} \quad X_{j_1} \leq X_{j_2} \leq \cdots \leq X_{j_n}, \tag{46b}$$

$$X_j = \sum_{i \in I_j} x_i, \quad \forall j \in N, \tag{46c}$$

$$\sum_{i \in M} x_i = K, \tag{46d}$$

$$x_i \in \{0, 1\}, \quad \forall i \in M. \tag{46e}$$

In Figure 9, a unique optimal solution has sensors at nodes 1 and 5 implying that nodes 2, 3, and 5 are monitored by both sensors while node 4 is monitored by only one sensor. Note that, in general, there may be multiple optimal solution. This problem can be solved by constructing counting functions as described in Section 2.4. However, whereas for the equitable facility location problem [42] the counting function for each location j is represented by a single constraint, here the representation of counting functions adds a large number of variables and constraints into the problem. Now, suppose that the probability of detecting an intruder at node j from a sensor at node i is $p_{ij} > 0$ for $i \in I_j$. Then the protection level provided to j is the probability that an intruder will be detected at node j by at least one sensor from among those placed in the set I_j. Although the formulation of this case is similar to the formulation above, the number of possible distinct outcomes can be much larger. As discussed in Section 2.4, this would necessitate employing a different solution method that is not based on counting functions (see [145]).

5. Complexity Issues

Essentially, fair optimization models are based on concave piecewise linear criteria possibly replacing a linear criterion of the total output maximization. Such criteria, implementable with auxiliary linear inequalities, in most cases do not significantly affect the complexity of the original optimization problems. In particular, problems represented by linear programming remain linear programs in their fair optimization versions (single LP problem in the case of fair OWA aggregations, or a sequence of LP problems for the MMF models). Certainly, some specialized algorithms taking into account the structure of the problem in hand can be more efficient than the general linear programming techniques. Consider resource allocation problem (23a)–(23d) in Section 2.4 which have a lexicographic maximin objective. For certain performance functions (e.g., linear and exponential functions) a lexicographic maximin solution is obtained by manipulations of closed-form expressions in a polynomial time. As presented in [5, Ch. 3.3], depending on the algorithm employed, the computational effort for solving (23a)–(23d) is $O(mn)$ or $O(m^2n)$ where m is the number of activities in the set I and n is the number of resources in the set J. Moreover, the same complexity is achieved for some content distribution problems in tree networks described in Section 3.6. With respect to communication networks applications, a well-known example of such a specialized algorithm is the already mentioned water-filling algorithm (see Section 3.2). Another example is a special case of the single source traffic allocation problem (also see Section 3.2), for which Megiddo [146] introduced a polynomial-time MMF algorithm which applies to splittable (fractional) flows. As presented in Section 2.4, there exist simple polynomial time techniques for solving general convex MMF problems. Thus, when applied to networks problems, the algorithms do not depend on any specific traffic routing problem formulation and is sufficiently general to be applied to a broad class of traffic routing and capacity allocation problems.

Generally, MMF optimization problems on convex attainable sets are characterized by polynomial complexity [92]. Polynomial algorithms may be developed for various specific forms of load balancing problems. For instance, in [147] a polynomial algorithm to determine the MMF optimal bandwidth allocation in order to satisfy the communication needs between two private networks. The algorithm is guaranteed to converge in finite number of steps, and for linear costs its complexity is $O(|\mathscr{V}|^5)$.

Nonconvex attainable sets usually results in NP-hard complexity of the corresponding fair optimization problems. In the network environment this is the case of single-path flows (unsplittable flows). In particular, a single-source multiple-sink demand MMF optimization of single-path flows in a directed network was proven NP-hard in [148]. Nilsson [149] generalized this result showing that general MMF unsplittable-flow problems on undirected networks are NP-hard. This applies to the case when each demand may use any path as well as to the case when each demand may use one path from a predefined list. Actually, it is proven there that in both cases obtaining just the first entry of the sorted allocation vector (the standard maximin) is NP-hard in itself. Observe that this shows that all corresponding fairness optimization models are NP-hard as they must take into account that criterion. Single-path optimization problems remain NP-hard also when fairness is implemented as a constraint rather than a criterion. Amaldi et al. [150] showed that the Max-Throughput Single-Path Network Routing subject to MMF flow allocation is NP-hard even with equal (unit) capacities for all links. Nilsson [149] has also shown than nonconvexity introduced by modular flows (granular) causes that even splittable traffic allocation problems become NP-hard. Therefore, there is an emerging need to develop approximate or heuristic algorithms for such problems. Early results in this area show that several communication network problems with PF or OWA fairness criteria can be effectively handled by meta-heuristic approaches [80, 106, 151].

In location and allocation problems the general fairness (equitable) models may be viewed as the so-called ordered median solution concepts, corresponding to the OWA criterion with monotonic weights. Such a criterion may be implemented with simple auxiliary linear inequalities. Nevertheless, even standard (median or center) multifacility location problems on general networks are usually NP-hard and the same remains valid for the ordered median problems. For tree networks, however, polynomial time algorithms exist. Dynamic programming algorithm for the ordered median problem presented in [152] has time complexity of $O(pm^8)$ for the general problem, and just $O(pm^4)$ for the node restricted problem. Polynomial algorithms exist also for the single facility location ordered median problems [153] with complexity $O(m^3\log^2 m)$ for trees and $O(|\mathscr{E}|m^3\log^2 m)$ for general networks.

In this survey we have not discussed in detail fair optimization in connection to problems which can be related to abstract networks or analyzed with some networks. An important wide group of such problems is related to job-shop scheduling. Most approaches for the job-shop scheduling problem deal with the makespan criterion, that is, the maximum completion time of all jobs. Still, there are various criteria that consider the due dates of jobs, and aim at minimizing the tardiness of jobs or the fact that jobs are late, that is, not completed before their due dates. Actually, simple aggregations of a number of such uniform criteria are commonly applied. Each single criterion applies to one scheduling object like job or affected agent, and a need for aggregations providing fairness arises. Note that any fair aggregation is strictly increasing, thus, satisfying the condition of the so-called regular scheduling criterion (i.e., it is an increasing function of the completion times of the jobs, i.e., it is always optimal to start and complete jobs as early as possible). The job-shop scheduling problems with regular criteria are well studied. For the nonpreemptive two machine job-shop scheduling problem with a fixed number of jobs any regular criterion can be solved in polynomial time [154]. Generally, the n-job m-machine job-shop problem belongs to the class of NP-hard problems [155–157] though there are exceptions for specific problems. Nevertheless, generic efficient approaches are available for approximate solving the job-shop scheduling problem with regular criteria [158]. Importance of fairness issues has been recently recognized in just-in time sequencing problems [13] in apportionment concepts [76, 159]. Very few fair optimization approaches have been presented to job-shop scheduling, although already in 1989 such approaches were considered in [160]. Specifically, a lexicographic minimax objective was analyzed for the production smoothness of multiple feeder shops that produce components for custom-made products assembled at a final assembly shop. Finally, we mention that the fair dominance rules were used in a multiobjective method to solve reentrant hybrid flow shop scheduling problem [161].

6. Concluding Remarks

In systems which serve many entities there is a need to respect some fairness rules. Extensive progress in fair optimization methodology, made in the last three decades, resulted in a variety of techniques enabling to generate fair and efficient solutions. In particular, allocation problems related to communication networks and location-allocation problems are the areas where the fair optimization concepts are extensively developed and widely applied. Within the networking applications the lexicographic maximin approach (or the related max-min fairness approach) is the most widely used. The recent book by Luss [5] exhibits a variety of models with a lexicographic maximin objective and the corresponding algorithms in the context of resource allocation. Many of these models apply to communication network and location problems. Since this approach may lead to significant losses in the overall efficiency (e.g., throughput of the network), the proportional fairness or other utility based approaches (like α-fairness) are also applied. In location-allocation problems the fairness understood as equity is usually quantified with inequality measures to be minimized or treated with minimax optimization, called center solution concept. The latter is applied especially for emergency facilities location and recently is considered with a lexicographically regularized criterion to lexicographic minimax. The inequality measures are scalar indices based on some measurement of the spread of outcomes. Direct minimization of the inequality measures contradicts the optimization of individual outcomes, but several inequality measures can be combined with the mean outcome into the equitable criteria, thus, allowing to generate various fair solutions.

A wide variety of fair optimization models and algorithms supporting efficient and fair allocation in complex systems has been introduced and studied in the literature. In most cases, they can be effectively used to generate various fair allocation schemes while taking into account the problem specificities. Nevertheless, problems with discrete structure lead to massive computations questioning possibility to achieve any fair solution in a reasonable time. Therefore, there is a need to develop approximate or heuristic algorithms for such problems.

Frequently, one may be interested in putting into allocation models some additional service importance weights. The importance weights are easily incorporated into the scalar inequality measures [59, 137, 162] or the Jain fairness index [163] as well as in proportional fairness [7]. There are also possibilities to introduce importance weights into the general fair preferences [164] and fair optimization models. In particular, the OWA aggregations (18) may be extended to the corresponding Weighted OWA (WOWA) aggregations [165, 166] which still remain LP computable [167, 168], while metaheuristic may be also applied [169]. The performance functions in a lexicographic minimax objective function may also include demand weights, cf. [31, 43, 170], and [5, Ch. 1].

Vector fair optimization approaches taking into account multiattribute outcomes are still underexplored. In resource allocation context this relates to problems with multiple types of resources where the users request different ratios of different resources. A typical example is datacenters processing jobs with heterogeneous resource requirements on CPU, memory, network bandwidth, and so forth. Recently proposed (vector) fairness measure [171], called dominant resource fairness, allocates resources according to max-min fairness on dominant resource shares. Köppen et al. [172] have extended the Jain fairness index [61] to the multiattribute case. By means of a leximin procedure, an allocation can be selected where the smallest among the Jain fairness indexes takes the largest value. This extends the notion of an allocation where fairness is achieved only for a single allocation metric. A unifying framework addressing the fairness-efficiency tradeoff in the light of multiple types of resources has been developed in [173].

Another still underexplored area of fair network optimization is related to distributed optimization process and related models [174]. In some equitable optimization problems, as shown in [113], the optimization algorithm can be implemented in a distributed mode where most of the computations are done independently and in parallel at the nodes. However, in most cases the distributed approaches to fairness must be based on game theory rather than on direct optimization [175–177].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

Research conducted by W. Ogryczak, M. Pióro, and A. Tomaszewski was supported by the National Science Centre (Poland) under Grant 2011/01/B/ST7/02967 "Integer programming models for joint optimization of link capacity assignment, transmission scheduling, and routing in fair multicommodity flow networks."

References

[1] T. Ibaraki and N. Katoh, *Resource Allocation Problems: Algorithmic Approaches*, MIT Press, Cambridge, Mass, USA, 1988.

[2] N. Katoh and T. Ibaraki, "Resource allocation problems," in *Handbook of Combinatorial Optimization*, D.-Z. Du and P. M. Pardalos, Eds., pp. 159–260, Kluwer Academic, Dordrecht, The Netherlands, 1998.

[3] A. P. Sage and C. D. Cuppan, "On the systems engineering and management of systems of systems and federations of systems," *Information, Knowledge, and System Management*, vol. 2, pp. 325–345, 2001.

[4] L. Wang, L. Fang, and K. W. Hipel, "Fairness in resource allocation in a system of systems," in *Proceedings of the IEEE*

International Conference on System of Systems Engineering (SoSE '07)*, pp. 1–6, San Antonio, Tex, USA, April 2007.

[5] H. Luss, *Equitable Resource Allocation: Models, Algorithms, and Applications*, John Wiley & Sons, Hoboken, NJ, USA, 2012.

[6] R. Denda, A. Banchs, and W. Effelsberg, "The fairness challenge in computer networks," in *Quality of Future Internet Services*, vol. 1922 of *Lecture Notes in Computer Science*, pp. 208–220, 2000.

[7] M. Pióro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, San Francisco, Calif, USA, 2004.

[8] A. D. Procaccia, "Cake cutting: not just child's play," *Communications of the ACM*, vol. 56, no. 7, pp. 78–87, 2013.

[9] W. Ogryczak, "Inequality measures and equitable approaches to location problems," *European Journal of Operational Research*, vol. 122, no. 2, pp. 374–391, 2000.

[10] L. Wang, L. Fang, and K. W. Hipel, "Basin-wide cooperative water resources allocation," *European Journal of Operational Research*, vol. 190, no. 3, pp. 798–817, 2008.

[11] J. Jaffe, "Bottleneck ow control," *IEEE Transactions on Communications*, vol. 7, pp. 207–237, 1980.

[12] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, Englewood Cliffs, NJ, USA, 1987.

[13] W. Kubiak, *Proportional Optimization and Fairness*, Springer, New York, NY, USA, 2009.

[14] D. Bertsimas, G. Lulli, and A. Odoni, "An integer optimization approach to large-scale air traffic flow management," *Operations Research*, vol. 59, no. 1, pp. 211–227, 2011.

[15] R. S. Klein and H. Luss, "Minimax resource allocation with tree structured substitutable resources," *Operations Research*, vol. 39, pp. 285–295, 1991.

[16] R. S. Klein, H. Luss, and U. G. Rothblum, "Minimax resource allocation problems with resource-substitutions represented by graphs," *Operations Research*, vol. 41, no. 5, pp. 959–971, 1993.

[17] J. Pang and C. S. Yu, "A min-max resource allocation problem with substitutions," *European Journal of Operational Research*, vol. 41, no. 2, pp. 218–223, 1989.

[18] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, Englewood Cliffs, NJ, USA, 1993.

[19] J. Malczewski and W. Ogryczak, "The multiple criteria location problem: 1. A generalized network model and the set of efficient solutions," *Environment & Planning A*, vol. 27, no. 12, pp. 1931–1960, 1995.

[20] W. Ogryczak, K. Studziński, and K. Zorychta, "A solver for the multi-objective transshipment problem with facility location," *European Journal of Operational Research*, vol. 43, no. 1, pp. 53–64, 1989.

[21] M. Köppen, "Relational optimization and its application: from bottleneck flow control to wireless channel allocation," *Informatica*, vol. 24, no. 3, pp. 413–433, 2013.

[22] W. Ogryczak, "Comments on properties of the minmax solutions in goal programming," *European Journal of Operational Research*, vol. 132, no. 1, pp. 17–21, 2001.

[23] W. Ogryczak, "On goal programming formulations of the reference point method," *Journal of the Operational Research Society*, vol. 52, no. 6, pp. 691–698, 2001.

[24] W. Ogryczak, "Multicriteria models for fair resource allocation," *Control and Cybernetics*, vol. 36, no. 2, pp. 303–332, 2007.

[25] J. Rawls, "Justice as fairness," *Philosophical Review*, pp. 164–194, 1958.

[26] J. Rawls, *The Theory of Justice*, Harvard University Press, Cambridge, Mass, USA, 1971.

[27] M. Dresher, *Games of Strategy: Theory and Applications*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1961.

[28] D. Schmeidler, "The nucleolus of a characteristic function game," *SIAM Journal on Applied Mathematics*, vol. 17, pp. 1163–1170, 1969.

[29] J. A. M. Potters and S. H. Tijs, "The nucleolus of a matrix game and other nucleoli," *Mathematics of Operations Research*, vol. 17, no. 1, pp. 164–174, 1992.

[30] J. R. Rice, "Tchebycheff approximation in a compact metric space," *Bulletin of the American Mathematical Society*, vol. 68, pp. 405–410, 1962.

[31] H. Luss and D. R. Smith, "Resource allocation among competing activities: a lexicographic minimax approach," *Operations Research Letters*, vol. 5, no. 5, pp. 227–231, 1986.

[32] F. A. Behringer, "A simplex based algorithm for the lexicographically extended linear maxmin problem," *European Journal of Operational Research*, vol. 7, no. 3, pp. 274–283, 1981.

[33] E. Marchi and J. A. Oviedo, "Lexicographic optimality in the multiple objective linear programming: the nucleolar solution," *European Journal of Operational Research*, vol. 57, no. 3, pp. 355–359, 1992.

[34] J. Lesca and P. Perny, "LP solvable models for multiagent fair allocation problems," in *Proceedings of the 19th European Conference on Artificial Intelligence*, vol. 215, pp. 393–398, IOS Press, Lisbon, Portugal, 2010.

[35] L. M. Betts, J. R. Brown, and H. Luss, "Minimax resource allocation problems with ordering constraints," *Naval Research Logistics*, vol. 41, no. 6, pp. 719–738, 1994.

[36] J. H. King, "Allocation of scarce resources in manufacturing facilities," *AT&T Technical Journal*, vol. 68, no. 3, pp. 103–113, 1989.

[37] R. S. Klein, H. Luss, and D. R. Smith, "A lexicographic minimax algorithm for multiperiod resource allocation," *Mathematical Programming*, vol. 55, no. 2, pp. 213–234, 1992.

[38] R. S. Klein, H. Luss, and U. G. Rothblum, "Multiperiod allocation of substitutable resources," *European Journal of Operational Research*, vol. 85, no. 3, pp. 488–503, 1995.

[39] Q. C. Nguyen and R. E. Stone, "A multiperiod minimax resource allocation problem with substitutable resources," *Management Science*, vol. 39, pp. 964–974, 1993.

[40] R. E. Burkard and F. Rendl, "Lexicographic bottleneck problems," *Operations Research Letters*, vol. 10, no. 5, pp. 303–308, 1991.

[41] F. Della Croce, V. T. Paschos, and A. Tsoukias, "An improved general procedure for lexicographic bottleneck problems," *Operations Research Letters*, vol. 24, no. 4, pp. 187–194, 1999.

[42] W. Ogryczak, "On the lexicographic minimax approach to location problems," *European Journal of Operational Research*, vol. 100, no. 3, pp. 566–585, 1997.

[43] H. Luss, "On equitable resource allocation problems: a lexicographic minimax approach," *Operations Research*, vol. 47, no. 3, pp. 361–378, 1999.

[44] W. Ogryczak, M. Pioro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problem," *Journal of Telecommunications and Information Technology*, vol. 2005, no. 3, pp. 43–56, 2005.

[45] B. Radunović and J.-Y. Le Boudec, "A unified framework for max-min and min-max fairness with applications," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1073–1083, 2007.

[46] D. Nace and M. Pióro, "Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.

[47] R. M. Salles and J. A. Barria, "Lexicographic maximin optimisation for fair bandwidth allocation in computer networks," *European Journal of Operational Research*, vol. 185, no. 2, pp. 778–794, 2008.

[48] W. Ogryczak and T. Śliwiński, "On direct methods for lexicographic min-max optimization," in *Computational Science and Its Applications—ICCSA 2006*, vol. 3982 of *Lecture Notes in Computer Science*, pp. 802–811, 2006.

[49] H. Dalton, "The measurement of the inequality of income," *The Measurement of the Inequality of Income*, vol. 30, pp. 348–361, 1920.

[50] A. C. Pigou, *Wealth and Welfare*, Macmillan, London, UK, 1912.

[51] J. Konow, "A positive theory of economic fairness," *Journal of Economic Behavior and Organization*, vol. 31, no. 1, pp. 13–35, 1996.

[52] J. Konow, "Which is the fairest one of all? A positive analysis of justice theories," *Journal of Economic Literature*, vol. 41, no. 4, pp. 1188–1239, 2003.

[53] S. J. Brams and A. D. Taylor, *Fair Division: From Cake Cutting to Dispute Resolution*, Cambridge University Press, New York, NY, USA, 1996.

[54] H. Steinhaus, "Sur la division pragmatique," *Econometrica*, vol. 17, pp. 315–319, 1949.

[55] A. B. Atkinson, "On the measurement of inequality," *Journal of Economic Theory*, vol. 2, pp. 244–263, 1970.

[56] M. Rothschild and J. E. Stiglitz, "Some further results on the measurement of inequality," *Journal of Economic Theory*, vol. 6, no. 2, pp. 188–204, 1973.

[57] A. Sen, *On Economic Inequality*, Clarendon Press, Oxford, UK, 1973.

[58] M. T. Marsh and D. A. Schilling, "Equity measurement in facility location analysis: a review and framework," *European Journal of Operational Research*, vol. 74, no. 1, pp. 1–17, 1994.

[59] M. C. López-de-los-Mozos and J. A. Mesa, "The sum of absolute differences on a network: algorithm and comparison with other equality measures," *INFOR Journal*, vol. 41, no. 2, pp. 195–210, 2003.

[60] J. A. Mesa, J. Puerto, and A. Tamir, "Improved algorithms for several network location problems with equality measures," *Discrete Applied Mathematics*, vol. 130, no. 3, pp. 437–448, 2003.

[61] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer system," Tech. Rep., Eastern Research Laboratory, Digital Equipment Corp., 1984.

[62] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM '10)*, pp. 1–9, IEEE, San Diego, Calif, USA, March 2010.

[63] M. M. Kostreva and W. Ogryczak, "Linear optimization with multiple equitable criteria," *RAIRO Operations Research*, vol. 33, no. 3, pp. 275–297, 1999.

[64] M. M. Kostreva, W. Ogryczak, and A. Wierzbicki, "Equitable aggregations and multiple criteria analysis," *European Journal of Operational Research*, vol. 158, no. 2, pp. 362–377, 2004.

[65] W. Ogryczak, "Fair optimization: methodological foundations of fairness in network resource allocation," in *Proceedings of the IEEE 38th Annual Computer Software and Applications Conference Workshops (COMPSACW '14)*, pp. 43–48, IEEE, 2014.

[66] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, vol. 143 of *Mathematics in Science and Engineering*, Academic Press, New York, NY, USA, 1979.

[67] F. P. Kelly, A. K. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 206–217, 1997.

[68] M. Pióro, G. Malicsko, and G. Fodor, "Optimal link capacity dimensioning in proportionally fair networks," in *NETWORKING 2002: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*, vol. 2345 of *Lecture Notes in Computer Science*, pp. 277–288, Springer, Berlin, Germany, 2002.

[69] J. Nash, "The bargaining problem," *Econometrica*, vol. 18, pp. 155–162, 1950.

[70] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.

[71] C. Y. Lee and H. K. Cho, "Discrete bandwidth allocation considering fairness and transmission load in multicast networks," *Computers and Operations Research*, vol. 34, no. 3, pp. 884–899, 2007.

[72] C. Y. Lee, Y. P. Moon, and Y. J. Cho, "A lexicographically fair allocation of discrete bandwidth for multirate multicast traffics," *Computers & Operations Research*, vol. 31, no. 14, pp. 2349–2363, 2004.

[73] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," *Operations Research*, vol. 59, no. 1, pp. 17–31, 2011.

[74] D. Bertsimas, V. F. Farias, and N. Trichakis, "On the efficiency-fairness trade-off," *Management Science*, vol. 58, no. 12, pp. 2234–2250, 2012.

[75] A. Tang, J. Wang, and S. H. Low, "Is fair allocation always inefficient," in *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, pp. 35–45, Hong Kong, China, 2004.

[76] H. P. Young, *Equity in Theory and Practice*, Princeton University Press, Princeton, NJ, USA, 1994.

[77] H. Moulin, *Axioms of Cooperative Decision Making*, Cambridge University Press, New York, NY, USA, 1988.

[78] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[79] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*, Wiley, New York, NY, USA, 1994.

[80] M. Köppen, K. Yoshida, R. Verschae, M. Tsuru, and Y. Oie, "Comparative study on meta-heuristics for achieving parabolic fairness in wireless channel allocation," in *Proceedings of the IEEE/IPSJ 12th International Symposium on Applications and the Internet (SAINT '12)*, pp. 302–307, IEEE, July 2012.

[81] R. R. Yager, "On the analytic representation of the Leximin ordering and its application to flexible constraint propagation," *European Journal of Operational Research*, vol. 102, no. 1, pp. 176–192, 1997.

[82] M. Majdan and W. Ogryczak, "Determining OWA operator weights by mean absolute deviation minimization," in *Artificial Intelligence and Soft Computing*, vol. 7267 of *Lecture Notes in Computer Science*, pp. 283–291, Springer, 2012.

[83] Z. Xu, "An overview of methods for determining OWA weights," *International Journal of Intelligent Systems*, vol. 20, no. 8, pp. 843–865, 2005.

[84] W. Ogryczak, "Tail mean and related robust solution concepts," *International Journal of Systems Science: Principles and Applications of Systems and Integration*, vol. 45, no. 1, pp. 29–38, 2014.

[85] W. Ogryczak and A. Tamir, "Minimizing the sum of the *k* largest functions in linear time," *Information Processing Letters*, vol. 85, no. 3, pp. 117–122, 2003.

[86] W. Ogryczak and T. Śliwiński, "On solving linear programs with the ordered weighted averaging objective," *European Journal of Operational Research*, vol. 148, no. 1, pp. 80–91, 2003.

[87] R. R. Yager, "Constrained OWA aggregation," *Fuzzy Sets and Systems*, vol. 81, no. 1, pp. 89–101, 1996.

[88] W. Ogryczak and A. Ruszczynski, "Dual stochastic dominance and related mean-risk models," *SIAM Journal on Optimization*, vol. 13, no. 1, pp. 60–78, 2002.

[89] A. Müller and D. Stoyan, *Comparison Methods for Stochastic Models and Risks*, John Wiley & Sons, Chichester, UK, 2002.

[90] H. Luss, "An algorithm for separable nonlinear minimax problems," *Operations Research Letters*, vol. 6, no. 4, pp. 159–162, 1987.

[91] H. Luss, "A nonlinear minimax allocation problem with multiple knapsack constraints," *Operations Research Letters*, vol. 10, no. 4, pp. 183–187, 1991.

[92] A. Tomaszewski, "A polynomial algorithm for solving a general max-min fairness problem," *European Transactions on Telecommunications*, vol. 16, no. 3, pp. 233–240, 2005.

[93] W. Ogryczak and T. Śliwiński, "Sequential algorithms for max-min fair bandwidth allocation," in *Proceedings of the European Computing Conference*, V. Mastorakis, V. Mladenov, and V. T. Kontargyri, Eds., vol. 27 of *Lecture Notes in Electrical Engineering*, pp. 511–522, Springer, New York, NY, USA, 2009.

[94] W. Ogryczak and T. Śliwiński, "Sequential algorithms for exact and approximate max-min fair bandwidth allocation," in *Proceedings of the 15th International Telecommunications Network Strategy and Planning Symposium (NETWORKS '12)*, pp. 1–6, Rome, Italy, October 2012.

[95] W. Ogryczak, T. Sliwinski, and A. Wierzbicki, "Fair resource allocation schemes and network dimensioning problems," *Journal of Telecommunications and Information Technology*, vol. 2003, no. 3, pp. 34–42, 2003.

[96] W. Ogryczak, A. Wierzbicki, and M. Milewski, "A multi-criteria approach to fair and efficient bandwidth allocation," *Omega*, vol. 36, no. 3, pp. 451–463, 2008.

[97] D. Nace and J. B. Orlin, "Lexicographically minimum and maximum load linear programming problems," *Operations Research*, vol. 55, no. 1, pp. 182–187, 2007.

[98] P. Nilsson and M. Pióro, "Max-min fair distribution of modular network flows on fixed paths," in *Proceedings of the 5th International IFIP-TC6 Networking Conference*, vol. 3976 of *Lecture Notes in Computer Science*, pp. 916–927, Springer, Berlin, Germany, 2006.

[99] E. Amaldi, A. Capone, S. Coniglio, and L. G. Gianoli, "Network optimization problems subject to max-min fair flow allocation," *IEEE Communications Letters*, vol. 17, no. 7, pp. 1463–1466, 2013.

[100] D. Santos, A. de Sousa, F. Alvelos, M. Dzida, and M. Pióro, "Optimization of link load balancing in multiple spanning tree routing networks," *Telecommunication Systems*, vol. 48, no. 1-2, pp. 109–124, 2011.

[101] D. Santos, A. de Sousa, F. Alvelos, and M. Pióro, "Optimizing network load balancing: an hybridization approach of metaheuristics with column generation," *Telecommunication Systems*, vol. 52, no. 2, pp. 959–968, 2013.

[102] P. Å. Nilsson and M. Pióro, "Solving dimensioning tasks for proportionally fair networks carrying elastic traffic," *Performance Evaluation*, vol. 49, no. 1–4, pp. 371–386, 2002.

[103] E. Kubilinskas, P. Nilsson, and M. Pióro, "Design models for robust multi-layer next generation Internet core networks carrying elastic traffic," *Journal of Network and Systems Management*, vol. 13, no. 1, pp. 57–76, 2005.

[104] M. Pióro, E. Kubilinskas, and P. Nilsson, "Resilient dimensioning of proportionally fair networks," *European Transactions on Telecommunications*, vol. 16, no. 3, pp. 241–251, 2005.

[105] M. Pióro, M. Żotkiewicz, B. Staehle, D. Staehle, and D. Yuan, "On max-min fair flow optimization in wireless mesh networks," *Ad Hoc Networks*, vol. 13, pp. 134–152, 2014.

[106] J. Hurkala and T. Śliwiński, "Fair flow optimization with advanced aggregation operators in Wireless Mesh Networks," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS '12)*, pp. 415–421, IEEE, September 2012.

[107] A. P. Wierzbicki, M. Makowski, and J. Wessels, Eds., *Model Based Decision Support Methodology with Environmental Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.

[108] W. Ogryczak and A. Wierzbicki, "On multi-criteria approaches to bandwidth allocation," *Control and Cybernetics*, vol. 33, no. 3, pp. 427–448, 2004.

[109] W. Ogryczak, M. Milewski, and A. Wierzbicki, "On fair and efficient bandwidth allocation by the multiple target approach," in *Proceedings of the 2nd Conference on Next Generation Internet Design and Engineering (NGI '06)*, pp. 48–55, IEEE, April 2006.

[110] A. Tomaszewski, "Design of optical wireless networks with fair traffic flows," *Journal of Applied Mathematics*, vol. 2014, Article ID 938483, 5 pages, 2014.

[111] H. Luss, "Equitable bandwidth allocation in content distribution networks," *Naval Research Logistics*, vol. 57, no. 3, pp. 266–278, 2010.

[112] S. Sarkar and L. Tassiulas, "Distributed algorithms for computation of fair rates in multirate multicast trees," in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM '00)*, pp. 52–61, March 2000.

[113] H. Luss, "A distributed algorithm for equitable bandwidth allocation for content distribution in a tree network," *Journal of the Operational Research Society*, vol. 63, no. 4, pp. 460–469, 2012.

[114] H. Luss, "An equitable bandwidth allocation model for video-on-demand networks," *Networks and Spatial Economics*, vol. 8, no. 1, pp. 23–41, 2008.

[115] S. Sarkar and L. Tassiulas, "Fair allocation of utilities in multirate multicast networks: a framework for unifying diverse fairness objectives," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 931–944, 2002.

[116] S. Sarkar and L. Tassiulas, "Fair bandwidth allocation for multicasting in networks with discrete feasible set," *IEEE Transactions on Computers*, vol. 53, no. 7, pp. 785–797, 2004.

[117] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithms for congestion avoidance in computer networks," *Computer Networks and ISDN Systems*, vol. 17, no. 1, pp. 1–14, 1989.

[118] L. Massoulié and J. W. Roberts, "Bandwidth sharing and admission control for elastic traffic," *Telecommunication Systems*, vol. 15, no. 1-2, pp. 185–201, 2000.

[119] D. Nace, N. Doan, E. Gourdin, and B. Liau, "Computing optimal max-min fair resource allocation for elastic flows," *IEEE/ACM Transactions on Networking*, vol. 14, no. 6, pp. 1272–1281, 2006.

[120] D. Nace, L. Nhat Doan, O. Klopfenstein, and A. Bashllari, "Max-min fairness in multi-commodity flows," *Computers and Operations Research*, vol. 35, no. 2, pp. 557–573, 2008.

[121] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 163–173, 2005.

[122] K. Wakuda, S. Kasahara, Y. Takahashi, Y. Kure, and E. Itakura, "A packet scheduling algorithm for max-min fairness in multihop wireless LANs," *Computer Communications*, vol. 32, no. 13-14, pp. 1437–1444, 2009.

[123] T. Bonald and J. Roberts, "Scheduling network traffic, SIGMETRICS perform," *Evaluation Review*, vol. 34, no. 4, pp. 29–35, 2007.

[124] S. Lai and B. Ravindran, "Achieving max-min lifetime and fairness with rate allocation for data aggregation in sensor networks," *Ad Hoc Networks*, vol. 9, no. 5, pp. 821–834, 2011.

[125] A. Gogu, D. Nace, S. Chatterjea, and A. Dilo, "Max-min fair link quality in WSN based on SINR," *Journal of Applied Mathematics*, vol. 2014, Article ID 693212, 11 pages, 2014.

[126] P. B. Coulter, "Measuring the inequality of urban public services: a methodological discussion with applications," *Policy Studies Journal*, vol. 8, pp. 693–698, 1980.

[127] L. D. Mayhew and G. Leonardi, "Equity, efficiency, and accessibility in urban and regional health-care systems," *Environment and Planning A*, vol. 14, pp. 1479–1507, 1982.

[128] M. Labbe, D. Peeters, and J.-F. Thisse, "Location on networks," in *Handbook in Operations Research and Management Science: Network Routing*, M. O. Ball, T. L. Magnanti, C. L. Monma, and G. L. Nemhauser, Eds., pp. 551–624, North-Holland, Amsterdam, The Netherlands, 1996.

[129] J. Malczewski and W. Ogryczak, "An interactive approach to the central facility location problem: locating pediatric hospitals in Warsaw," *Geographical Analysis*, vol. 22, no. 3, pp. 244–258, 1990.

[130] W. Ogryczak, "On the distribution approach to location problems," *Computers and Industrial Engineering*, vol. 37, no. 3, pp. 595–612, 1999.

[131] M. M. Kostreva, W. Ogryczak, and D. W. Tonkyn, "Relocation problems arising in conservation biology," *Computers and Mathematics with Applications*, vol. 37, no. 4-5, pp. 135–150, 1999.

[132] O. Berman, "Mean-variance location problems," *Transportation Science*, vol. 24, no. 4, pp. 287–293, 1990.

[133] O. Maimon, "The variance equity measure in locational decision theory," *Annals of Operations Research*, vol. 6, no. 5, pp. 147–160, 1986.

[134] E. Erkut, "Inequality measures for location problems," *Location Science*, vol. 1, pp. 199–217, 1993.

[135] O. Maimon, "An algorithm for the Lorenz measure in locational decisions on trees," *Journal of Algorithms*, vol. 9, no. 4, pp. 583–596, 1988.

[136] M. B. Mandell, "Modelling effectiveness-equity trade-offs in public service delivery systems," *Management Science*, vol. 37, no. 4, pp. 467–482, 1991.

[137] W. Ogryczak, "Inequality measures and equitable locations," *Annals of Operations Research*, vol. 167, pp. 61–86, 2009.

[138] M. M. Kostreva and W. Ogryczak, "Equitable approaches to location problems," in *Spatial Multicriteria Decision Making and Analysis: A Geographic Informatio Sciences Approach*, J .C. Thill, Ed., pp. 103–126, Ashgate, London, UK, 1999.

[139] W. Ogryczak, "On cent-dians of general networks," *Location Science*, vol. 5, no. 1, pp. 15–28, 1997.

[140] P. J. Slater, "Centers to centroids in graphs," *Journal of Graph Theory*, vol. 2, no. 3, pp. 209–222, 1978.

[141] A. Tamir, "The k-centrum multi-facility location problem," *Discrete Applied Mathematics*, vol. 109, no. 3, pp. 293–307, 2001.

[142] W. Ogryczak and M. Zawadzki, "Conditional median: a parametric solution concept for location problems," *Annals of Operations Research*, vol. 110, pp. 167–181, 2002.

[143] W. Ogryczak and T. Sliwinski, "On equitable approaches to resource allocation problems: the conditional minimax solution," *Journal of Telecommunications and Information Technology*, vol. 2002, no. 3, pp. 40–48, 2002.

[144] S. Nickel and J. Puerto, *Location Theory: A Unified Approach*, Springer, Berlin, Germany, 2005.

[145] A. Neidhardt, H. Luss, and K. R. Krishnan, "Data fusion and optimal placement of fixed and mobile sensors," in *Proceedings of the 3rd IEEE Sensors Applications Symposium (SAS '08)*, pp. 128–133, Atlanta, Ga, USA, February 2008.

[146] N. Megiddo, "Optimal flows in networks with multiple sources and sinks," *Mathematical Programming*, vol. 7, no. 1, pp. 97–107, 1974.

[147] L. Georgiadis, P. Georgatsos, K. Floros, and S. Sartzetakis, "Lexicographically optimal balanced networks," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 2, pp. 689–698, Anchorage, AK, USA, April 2001.

[148] J. Kleinberg, Y. Rabani, and E. Tardos, "Fairness in routing and load balancing," *Journal of Computer and System Sciences*, vol. 63, no. 1, pp. 2–21, 2001.

[149] P. Nilsson, *Fairness in communication and computer network design [Ph.D. thesis]*, Lund University, Lund, Sweden, 2006.

[150] E. Amaldi, S. Coniglio, L. G. Gianoli, and C. U. Ileri, "On single-path network routing subject to max-min fair flow allocation," *Electronic Notes in Discrete Mathematics*, vol. 41, no. 5, pp. 543–550, 2013.

[151] M. Köppen, K. Yoshida, K. Ohnishi, and M. Tsuru, "Meta-heuristic approach to proportional fairness," *Evolutionary Intelligence*, vol. 5, no. 4, pp. 231–244, 2012.

[152] J. Kalcsics, "Improved complexity results for several multifacility location problems on trees," *Annals of Operations Research*, vol. 191, pp. 23–36, 2011.

[153] J. Kalcsics, S. Nickel, J. Puerto, and A. Tamir, "Algorithmic results for ordered median problems," *Operations Research Letters*, vol. 30, no. 3, pp. 149–158, 2002.

[154] P. Brucker, S. A. Kravchenko, and Y. N. Sotskov, "On the complexity of two machine job-shop scheduling with regular objective functions," *OR Spektrum*, vol. 19, no. 1, pp. 5–10, 1997.

[155] P. Brucker, Y. N. Sotskov, and F. Werner, "Complexity of shop-scheduling problems with fixed number of jobs: a survey," *Mathematical Methods of Operations Research*, vol. 65, no. 3, pp. 461–481, 2007.

[156] M. R. Garey, D. S. Johnson, and R. Sethi, "The complexity of flowshop and jobshop scheduling," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 117–129, 1976.

[157] N. V. Shakhlevich, Y. N. Sotskov, and F. Werner, "Complexity of mixed shop scheduling problems: a survey," *European Journal of Operational Research*, vol. 120, no. 2, pp. 343–351, 2000.

[158] Y. Mati, S. Dauzère-Pérès, and C. Lahlou, "A general approach for optimizing regular criteria in the job-shop scheduling problem," *European Journal of Operational Research*, vol. 212, no. 1, pp. 33–42, 2011.

[159] M. L. Balinski and H. P. Young, *Fair Representation: Meeting the Ideal of One Man*, Yale University Press, New Haven, Conn, USA, 1982.

[160] H. Groeflin, H. Luss, M. B. Rosenwein, and E. T. Wahls, "Final assembly sequencing for just-in-time manufacturing," *International Journal of Production Research*, vol. 27, no. 2, pp. 199–213, 1989.

[161] F. Dugardin, F. Yalaoui, and L. Amodeo, "New multi-objective method to solve reentrant hybrid flow shop scheduling problem," *European Journal of Operational Research*, vol. 203, no. 1, pp. 22–31, 2010.

[162] M. C. López-de-los-Mozos, J. A. Mesa, and J. Puerto, "A generalized model of equality measures in network location problems," *Computers and Operations Research*, vol. 35, no. 3, pp. 651–660, 2008.

[163] Z. Chen, "A modified resource distribution fairness measure," *Journal of Mathematics*, vol. 2013, Article ID 956039, 6 pages, 2013.

[164] W. Ogryczak, "On principles of fair resource allocation for importance weighted agents," in *Proceedings of the International Workshop on Social Informatics (SOCINFO '09)*, pp. 57–62, Warsaw, Poland, June 2009.

[165] V. Torra, "The weighted OWA operator," *International Journal of Intelligent Systems*, vol. 12, no. 2, pp. 153–166, 1997.

[166] V. Torra and Y. Narukawa, *Modeling Decisions Information Fusion and Aggregation Operators*, Springer, Berlin, Germany, 2007.

[167] W. Ogryczak and T. Śliwiński, "On Optimization of the importance weighted OWA aggregation of multiple criteria," in *Computational Science and Its Applications—ICCSA 2007*, vol. 4705 of *Lecture Notes in Computer Science*, pp. 804–817, 2007.

[168] W. Ogryczak and T. Śliwiński, "On efficient WOWA optimization for decision support under risk," *International Journal of Approximate Reasoning*, vol. 50, no. 6, pp. 915–928, 2009.

[169] J. Hurkała and T. Śliwiński, "Threshold accepting heuristic for fair flow optimization in wireless mesh networks," *Journal of Applied Mathematics*, vol. 2014, Article ID 108673, 11 pages, 2014.

[170] Ľ. Buzna, M. Koháni, and J. Janáček, "An approximation aLgorithm for the facility location problem with lexicographic minimax objective," *Journal of Applied Mathematics*, vol. 2014, Article ID 562373, 12 pages, 2014.

[171] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica, "Dominant resource fairness: fair allocation of multiple resource types," in *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation*, pp. 24–37, USENIX Association, 2011.

[172] M. Köppen, K. Ohnishi, and M. Tsuru, "Multi-Jain fairness index of per-entity allocation features for fair and efficient allocation of network resources," in *Proceedings of the 5th International Conference on Intelligent Networking and Collaborative Systems (INCoS '13)*, pp. 841–846, Xi'an, China, 2013.

[173] C. Joe-Wong, S. Sen, T. Lan, and M. Chiang, "Multi-resource allocation: fairness-efficiency tradeoffs in a unifying framework," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 1206–1214, March 2012.

[174] A. Wierzbicki, *Trust and Fairness in Open, Distributed Systems*, Springer, Berlin, Germany, 2010.

[175] S. Hsu and A. S. Tsai, "A game-theoretic analysis of bandwidth allocation under a user-grouping constraint," *Journal of Applied Mathematics*, vol. 2013, Article ID 480962, 9 pages, 2013.

[176] N. Nisam, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds., *Algorithmic Game Theory*, Cambridge University Press, Cambridge, UK, 2007.

[177] K. Rzadca, D. Trystram, and A. Wierzbicki, "Fair game-theoretic resource management in dedicated grids," in *Proceedings of the 7th IEEE International Symposium on Cluster Computing and the Grid (CCGrid '07)*, pp. 343–350, May 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 725974, 12 pages
http://dx.doi.org/10.1155/2014/725974

Research Article

A Fairness Relation Based on the Asymmetric Choquet Integral and Its Application in Network Resource Allocation Problems

Aoi Honda and Mario Köppen

Kyushu Institute of Technology, 680-2 Kawazu, Iizuka, Fukuoka 820-8502, Japan

Correspondence should be addressed to Mario Köppen; mkoeppen@ieee.org

Received 6 March 2014; Revised 3 June 2014; Accepted 5 June 2014; Published 17 July 2014

Academic Editor: Wlodzimierz Ogryczak

The recent problem of network resource allocation is studied where pairs of users could be in a favourable situation, given that the allocation scheme is refined by some add-on technology. The general question here is whether the additional effort can be effective with regard to the user's experience of fairness. The computational approach proposed in this paper to handle this question is based on the framework of relational optimization. For representing different weightings for different pairs of users, the use of a fuzzy measure appears to be reasonable. The generalized Choquet integrals are discussed from the viewpoint of representing fairness and it is concluded that the asymmetric Choquet integral is the most suitable approach. A binary relation using the asymmetric Choquet integral is proposed. In case of a supermodular fuzzy measure, this is a transitive and cycle-free relation. The price of fairness with regard to a wireless channel allocation problem taking channel interference into account is experimentally studied and it can be seen that the asymmetric on relation actually selects allocations that perform on average between maxmin fairness and proportional fairness, and being more close to maxmin fairness as long as channel interference is not high.

1. Introduction

The rapid spread of wireless communication poses many challenges to the underlying networking technology and infrastructure. Daily experience of using wireless access teaches us the increased efficiency of provider solutions. At this stage of development, the major technical demand is the *efficient* reuse of existing resources or their usage expansion based on cost-efficient technical add-ons. That brings that the main valuation criterion today is the total utility of network infrastructure employment, with a lower focus on fairness aspects of resource allocation and distribution. For now, the standard user of a wireless infrastructure does not have the immediate experience of fairness, taking the access vector of other users into account—the needed information is simply not available for the user. This might change in the future and fairness might become the primary not the secondary aspect in comparison to efficiency. Recently, the relation between fairness and efficiency (in the sense of the total sum of per-user valuations) has become a relevant research issue. Obviously, a fair allocation is not an "optimal" allocation since the total valuation will be lower than the maximum possible. However, bounds do exist and in basic settings it is already known that the "price of fairness" is actually not so high [1].

In the specification of fairness, the so-called "standards of comparison" (SoC) give an important formal instrument for the specification of fair solutions. It refers to a modality of comparing two solutions, where a solution A passes a specific test with regard to another solution B. A classical example is the Nash standard of comparison for bargaining negotiations [2]; a solution is considered *better* than another solution if the relative wins are not outperformed by relative losses. If a solution A appears to be better than any other solution, that is, the relative losses are always larger than or equal to the relative wins, the solution is seen as a stable point in negotiations, and there is no incentive to deviate from this solution. Later on, a few issues with the Nash SoC led to the formulation of an alternative Kalai-Smorodinsky SoC [3]. This approach promotes the user that receives the least allocation (which is also the base for the economical and political ideas of Rawls [4]). Both approaches, originally developed in economical science, have been rediscovered in network telecommunication theory and are very popular approaches in current research; the Nash SoC is now better known as proportional fairness [5] and the Kalai-Smorodinsky SoC as maxmin fairness [6]. It comes out that, even historically, the cultural distinction between these two fundamental kinds of fairness existed for long time [7]. In between, a mediating approach that links proportional fairness and maxmin fairness, the so-called α-fairness, has been formulated as well [8]; if $\alpha = 0$ the relation refers to proportional fairness, and (with some caution) if $\alpha \to \infty$, the relation approximates maxmin fairness.

However, the fairness models so far are based on a per-user valuation of "wins" and "losses" where only a single user can decide what is in her interest and what not. For example, in proportional fairness, the SoC is based on the comparison between two allocation vectors x and y of the same dimension:

$$\sum_i \frac{y_i - x_i}{x_i} \leq 0. \quad (1)$$

One can see that each user, formally, contributes a single index value to the comparison test. However, as mentioned above for the case of wireless communication, increased demand of reuse of existing resources by new technical means and add-ons puts also *groups of users* into different situations that nevertheless need to be compared. There are many refined technologies explored these days that exhibit this property. One example is channel interference; in a wireless infrastructure, a base station BS has to allocate transmission channels (i.e., ranges of frequency bandwidth) to its subscriber stations (SS). The safe way is to allocate at most one SS to a channel. However, if SSs are well separated, that is, the channel interference of two SSs using the same frequency band is low, the same channel can be assigned to more than one SS. The question is how to know about this opportunity by proper measurement and about the best way to employ this. Several studies have been devoted to this technical challenge, but what they all have in common is the establishment of a pairwise relation between users, usually also weighted, as, for example, in [9–14].

Besides channel interference, there are many other newer networking technologies where this aspect of pairwise user weight appears; in cognitive radio, unused channel capacities of primary users can be on occasion allocated to secondary users. The pairwise relation here is represented by a conflict matrix that contains lower values where the chance that two users (one from the group of primary users, one from the group of secondary users) transmitting data in the same time intervall is low. The encounter probability also plays an important role in peer-to-peer networking, opportunistic networking, and vehicular networks.

Now, if considering above formalization of proportional fairness, how do we take this aspect into account; that is, the fact that a loss or win (relative or absolute) matters differently for different pairs of users? In other words, if, for example, user 1 and user 2 both experience losses when comparing two solutions could be taken differently from a situation where user 1 experiences a win and user 2 a loss. So far, no "pairwise standard of comparison" is known and it does not seem to be reasonable to define a different SoC for each different situation due to combinatorial explosion. The related problematic is about the price of fairness here; if a provider will consider a fair solution instead of the most efficient one, and taking the collision matrix information of a specific situation between users into account, is it worthy of the effort to implement the resource-reusing solution at all (i.e., to employ low channel interference, or to set up a schedule for cognitive radio secondary user channel assignment) if at the end no other allocation will be provided than without the technical add-on?

The primary goal of this paper is to provide a formal fairness approach that can distinguish groups of users and weight them differently where needed. For this, it is needed to understand the SoC in terms of (set-theoretic) relations, following [15]. Then, a specific relation will be introduced based on fuzzy measure theory and related fuzzy integrals. It happens that this relation comes out to be a transitive relation (allowing for ranking of solutions, as well as fast search algorithms) and can be used to valuate the efficiency losses related to making fair allocations.

Fuzzy measures [16, 17], which are called nonadditive measures [18] or capacity [19], are monotone and usually nonnegative set function. Since they can be used to express interactions between items that are not expressible by additive measures, many studies have been done on their application in fields such as subjective evaluation problems, decision-making support, and pattern recognition. The generalization with respect to fuzzy measures, a generalization of the Lebesgue integral, is proposed by several authors. The Choquet integral defined by Choquet [19] takes its place as one of them and is most widely used. The original Choquet integral is only for nonnegative functions, so that several generalizations for general functions which are not necessarily nonnegative have been given. There are mainly two types of generalization, one of which is called the symmetric Choquet integral or the Šipoš integral [20], and another one which is called the asymmetric Choquet integral is given in [18]. Concerning the details of these integrals, the first attempt was made by Denneberg [18]. This has been followed by Grabisch et al. They study properties and application to multicriteria decision making and so on, for example, in [21, 22]. Recent generalizations of the symmetric Choquet integral, employing averages and balanced weights over permutations, are the fusion Choquet integral [23] and the balancing Choquet integral [24].

On the other hand, preference modelling and the related preference prediction have become research fields of increasing importance. The use of the Choquet integral as a base for a preference relation has been the topic of [25] where it is used to solve combinatorial optimization problems. The application of a preference relation model based on Choquet integral in multiobjective dynamic programming is the topic of [26]. In these works, the optimality is prespecified and given by independent means. Here, we shall focus on the formal representations of fairness in resource distribution as optimality issue by itself. Our proposal is to evaluate distributions from the aspect of fairness.

In this paper, we compare the generalized Choquet integral, the symmetric Choquet integral, and the asymmetric Choquet integral from the viewpoint of representing fairness with the result that the asymmetric Choquet integral is more suitable for describing fairness. We also propose a fairness relation using the asymmetric Choquet integral.

The paper is organized as follows. Section 2 introduces notations and necessary materials about fuzzy measure and the Choquet integral and recalls the relational approach to fairness. Then, Section 3 presents discussion about two generalized Choquet integrals. The fairness relation is proposed in Section 4, and we give a sufficient condition that this relation is consistent and transitive. Section 5 gives a demonstration of how our fairness relation is used for decision making. Section 6 compares the solutions obtained by this fairness relation with the proportional and maxmin fair solutions by way of experiments.

2. Preliminaries

2.1. Relational Optimization. Speaking about optimality is speaking about comparing things. The basic mathematical instrument for the representation of optimality is the concept of a binary relation. Here, we want to recall some basic terms and definitions; for more details, see, for example, [15]. A (set-theoretic) binary relation R over a domain A is a subset of $A\times A$; that is, a set of ordered pairs (x, y) with $x, y \in A$. It is said that x is in relation R to y (sometimes also written as xRy or, if the focus is on comparison, $x\geq_R y$). There are alternative ways to represent relations, like a set comprehension, a mapping or function from A into the powerset of A, an incidence matrix or a directed graph. Also, relations can have properties. Two properties are of particular interest in the following; a relation is *symmetric* if from $(x, y) \in R$ it follows that also $(y, x) \in R$ and it is asymmetric if from $(x, y) \in R$ it always follows that (y, x) does not belong to R. Each relation can be decomposed into a symmetric and an asymmetric part, written as $R = I(R)\cup P(R)$, where $I(R)$ is a symmetric relation and $P(R)$ is an asymmetric relation. It is easy to decide to which part a pair $(x, y) \in R$ belongs in this decomposition; if (y, x) is also in R, it belongs to the symmetric part $I(R)$; if not, it belongs to the asymmetric part $P(R)$. It means that each relation, no matter what its domain is made of, shows mixed aspects of equality or similarity and a means of "betterness" or preference. For example, the real-valued "$\geq$" relation is decomposed into an equivalence relation "=" as symmetric part and a strictly larger relation ">" as asymmetric part (which is well symbolized by $\geq$). If we write $x\geq_R y$ for a relation, then $x>_R y$ will denote the asymmetric part, reflecting the "betterness" aspect of a relation R over domain A.

Any relation R can expose specific elements of its domain A which are commonly called greatest, maximal, minimal, and least elements. A *greatest* element x^* of the domain A with regard to R has the property that for any $x \in A$ (including x^*) it holds that $x^*\geq_R x$. All greatest elements comprise the *best set*. A *maximal* element x^* has the property that there is no $x \in P(R)$ such that $x>_R x^*$ and all such x^* comprise the *maximum set* (of A with respect to R). Minimal and least elements are defined correspondingly. Note that they do not need to exist in general.

One can easily see that a standard of comparison, as it was mentioned in the introduction section, is the same as the concept of a greatest element. However, often it is the case that best sets are empty. Therefore, it is convenient to refer to the maximum set as optimization goal, for several reasons. (1) In case of a finite domain, the maximum set is nonempty if the asymmetric part of the relation is cycle-free (i.e., the corresponding directed graph does not contain any cycles). (2) If there is a greatest element, it will also belong to the maximum set. (3) The maximum set can be seen as a "frontier"; they are not greatest elements but at least no better one is known.

In [15], the special case of fairness relations has been intensively discussed. It is comparable and easy to represent proportional fairness, maxmin fairness, leximin fairness, α-fairness, and so forth by this relational framework. Then, (Pareto-) efficient solutions can be found by finding the maximum sets for these relations. But, the advantage is also that the transition of a relational concept to another domain becomes possible (by applying the same formal definition), or that the relation can be specified in such a way that completely new aspects are taken into account (e.g., multiresource usage, collaboration between agents, or vector-valued evaluations). This is because the concept of maximality applies to any relation, no matter what its domain is. Therefore, if we want to handle the above mentioned problem of pairwise weights, we just have to specify a corresponding relation (that is preferably cycle-free and Pareto-efficient).

The last point that should be mentioned here is that the finding of maximum sets can be a challenging task, even if the complexity $O(n^2)$ is not so high. Thus, exhaustive search by a complete pairwise comparison can rapidly become intractable (e.g., a maximum set over a domain with one million elements would need 10^{12} pairwise comparisons). One possibility is to use metaheuristic search algorithms to approximate maximum sets; other alternatives are currently under investigations. In case the relation is transitive, the search for maximum sets can be performed in "batches"; that is, decompose the domain into nonoverlapping subsets and find the maximum set for each subset and then the maximum set of the union of the found maximum sets.

2.2. Computational Fairness. We have now used the term "fairness" several times without providing a general definition. For example, considering a given relation R, when is it safe to say that this is a fairness relation? Actually, there is no common understanding of fairness to which all researchers and practitioners would fully agree. It would go beyond the scope of this paper to consider the numerous considerations that have been done in order to provide a computational model of fairness, but at least we can summarize the basic ideas. The additional restriction here is that the focus is on the modality of a distribution, thus excluding other societal important concepts like procedural fairness or interaction fairness.

The common theme of all approaches to (distributional) fairness is to divide a set of resources among a number of agents. As already mentioned in the introduction, with far longest cultural history, we can find ideas about proportionality in the assignment (proportional according to demand, ability, suitability, availability, most available good, etc.) or the resolution of conflicts about contended goods in following an equity principle. Besides a relation point of view, in order to compare different possible assignments, the aspect can be also seen in aggregating the distributional aspects of all agents and thus keeping the spread of allocation differences small. This idea becomes strongly related to the theory of majorization [27], where numerical comparison between different ways of distributing the same total is strongly related to the Schur-convexity of the used comparison measure. The application of this approach and its various generalizations have been intensively studied, for example, in [28].

Any continuation of a formal approach to fairness (in the sense of a "definition of fairness") would need the specification of *axioms* that represent criteria derived from specific ways of how to balance—in a just way—the conflicts between agents that are in a better and agents that are in a worse situation—keeping in mind that fairness does generally not stand for the simple requirement of complete even allocations. Such axioms can refer to impartiality of distribution, consistency between solutions for different domains, mathematical assumptions about the domain, limiting behaviours, strategy-proofness of related procedures, or efficiency of the distribution.

Saying this, the considered aspect of fairness here is the balance between users who would gain more and users who would gain less in an alternative allocation. We will use aggregation operators for representing the joint state of all winning and all loosing agents, thus basically following the "recipe" of proportional fairness. But, as a new aspect, we also want to take a different weighting for groups of agents into account. Fuzzy integrals are a well-known formal approach for such a representation.

2.3. Fuzzy Measures and Integrals. We now want to recall basic issues of fuzzy measures and integrals in more detail. Most of the material can be gathered together from various publications and textbooks, but a comprehensive description, as it will be given now, might be of advantage for the understanding of the main proposal—a fairness relation based on the asymmetric Choquet integral for supermodular fuzzy measures.

Throughout this paper, we consider a finite universal set $N := \{1, 2, \ldots, n\}$ and 2^N denotes the power set of N.

Definition 1 (Fuzzy measure [16, 17]). A set function $\mu : 2^N \to [0, \infty)$ is a fuzzy measure if it satisfies the following conditions:

(i) $\mu(\emptyset) = 0$ and $\mu(N) < +\infty$,

(ii) for any $S, T \in 2^N$, $\mu(S) \le \mu(T)$ whenever $S \subset T$.

Properties (i) and (ii) express boundedness and monotonicity, respectively.

Definition 2 (superadditivity, subadditivity). A fuzzy measure μ is superadditive if μ satisfies $\mu(S \cup T) \ge \mu(S) + \mu(T)$ and subadditive if μ satisfies $\mu(S \cup T) \le \mu(S) + \mu(T)$, for any $S, T \in 2^N$ satisfying $S \cap T = \emptyset$.

Definition 3 (submodularity, supermodularity). A fuzzy measure μ is supermodular if μ satisfies $\mu(S \cup T) + \mu(S \cap T) \ge \mu(S) + \mu(T)$ and submodular if μ satisfies $\mu(S \cup T) + \mu(S \cap T) \le \mu(S) + \mu(T)$, for any $S, T \in 2^N$.

Definition 4 (dual measure). Let μ be a fuzzy measure. The dual measure of μ is defined by

$$m^d(S) := \mu(N) - \mu(S^c). \tag{2}$$

The dual measure of a fuzzy measure is also a fuzzy measure.

Proposition 5. *If μ is supermodular, then μ is superadditive. If μ is submodular, then μ is subadditive.*

Proof. When μ is supermodular, $\mu(S \cup T) + \mu(S \cap T) \ge \mu(S) + \mu(T)$ is satisfied for any $S, T \in 2^N$, and $S \cap T = \emptyset$. It implies $\mu(S \cup T) \ge \mu(S) + \mu(T)$ for any $S, T \in 2^N$ satisfying $S \cap T = \emptyset$. The second assertion is obtained in a similar manner. □

Proposition 6. *μ is supermodular if and only if μ^d is submodular. Similarly, μ is submodular if and only if μ^d is supermodular.*

Proof. Assume that μ is supermodular. Then, we have, for any $S, T \in 2^N$,

$$\begin{aligned} &\mu^d(S \cup T) + \mu^d(S \cap T) \\ &\quad = \mu(N) - \mu((S \cup T)^c) + \mu(N) - \mu((S \cap T)^c) \\ &\quad = 2\mu(N) - \mu(S^c \cap T^c) - \mu(S^c \cup T^c) \\ &\quad \le 2\mu(N) - \mu(S^c) - \mu(T^c) \\ &\quad = \mu^d(S) + \mu^d(T). \end{aligned} \tag{3}$$

Following the above in reverse, we obtain the converse. Replacing $\le$ with $\ge$, we obtain the second assertion. □

Proposition 7. *If μ is a superadditive fuzzy measure, then $\mu^d(S) \ge \mu(S)$ for any $S \in 2^N$.*

Proof. We have $\mu(S) + \mu(S^c) \le \mu(N)$ by superadditivity of μ. Hence, $\mu(S) \le \mu(N) - \mu(S^c) = \mu^d(S)$. □

Definition 8 (λ-fuzzy measure). A fuzzy measure μ is a λ-fuzzy measure if there exists $\lambda > -1$ such that $\mu(S \cup T) = \mu(S) + \mu(T) + \lambda\mu(S)\mu(T)$ for any $S, T \in 2^N$ satisfying $S \cap T = \emptyset$ and $\mu(N) = 1$.

Definition 9 (χ-fuzzy measure). A fuzzy measure μ on finite set $N = \{1, 2, \ldots, n\}$ is a χ-fuzzy measure if there exists

$$\chi \geq 1 - \frac{\min_{i\in N}\mu(\{i\})}{\sum_{i\in N}\mu(\{i\})} \tag{4}$$

such that

$$\mu(S) = \chi^{|S|-1}\sum_{i\in S}\mu(\{i\}) \tag{5}$$

for any $S \in 2^N$. Here, $|S|$ denotes the cardinal number of S.

$\lambda \geq 0$ and $\chi \geq 1$ imply superadditivity, and $\lambda \leq 0$ and $\chi \leq 1$ imply subadditivity. The dual measure of a λ-fuzzy measure and of a χ-fuzzy measure are not necessarily a λ-fuzzy measure and a χ-fuzzy measure.

Proposition 10. *Let μ be a λ-fuzzy measure. Then, μ is superadditive if and only if μ is supermodular. Similarly, μ is subadditive if and only if μ is submodular.*

Proof. The sufficiencies hold by Proposition 5. We show the necessities.

Assume μ is a superadditive λ-fuzzy measure; that is, $\lambda \geq 0$. Then, we have, for any $S, T \in 2^N$,

$$\begin{aligned}
&\mu(S\cup T) + \mu(S\cap T)\\
&\quad = \mu(S) + \mu(T\setminus S) + \lambda\mu(S)\mu(T\setminus S) + \mu(S\cap T)\\
&\quad \geq \mu(S) + \mu(T\setminus S) + \lambda\mu(S\cap T)\mu(T\setminus S) + \mu(S\cap T)\\
&\quad = \mu(S) + \mu(T).
\end{aligned} \tag{6}$$

Replacing $\leq$ with $\geq$, we obtain the second assertion. □

Definition 11 (Choquet integral [19]). Given a fuzzy measure $\mu : 2^N \to [0,1]$ and a nonnegative function f on N, the Choquet integral of f with respect to μ is defined by

$$(C)\int f\,d\mu = \int_0^{+\infty}\mu(\{xf(x) > r\})\,dr. \tag{7}$$

In the case that N is a finite set, (7) can be rewritten as

$$(C)\int f\,d\mu := \sum_{i=1}^{n}(a_i - a_{i-1})\,\mu(S_i), \tag{8}$$

where $a_i := f(i)$, $a_0 = 0$, $a_1 \leq a_2 \leq \cdots \leq a_n$, and $S_i := \{i, \ldots, n\}$. (cf. Figure 1). Note that this computation will include a permutation reordering of the a_i to ensure that they are in nondecreasing order.

The Choquet integral can be generalized to general functions, which are not necessarily nonnegative, in several ways. Here, we show the following two generalizations, the asymmetric Choquet integral and the symmetric Choquet integral.

Definition 12 (asymmetric Choquet integral [18]). Given a fuzzy measure $\mu : 2^N \to [0,1]$ and a function f on N, the asymmetric Choquet integral of f with respect to μ is defined by

$$\begin{aligned}
(\mathrm{ASC})\int f\,d\mu := &\int_{-\infty}^{0}(\mu(\{x \mid f(x) > r\})\\
&\qquad -m(N))\,dr\\
&+\int_0^{+\infty}\mu(x \mid f(x) > r)\,dr.
\end{aligned} \tag{9}$$

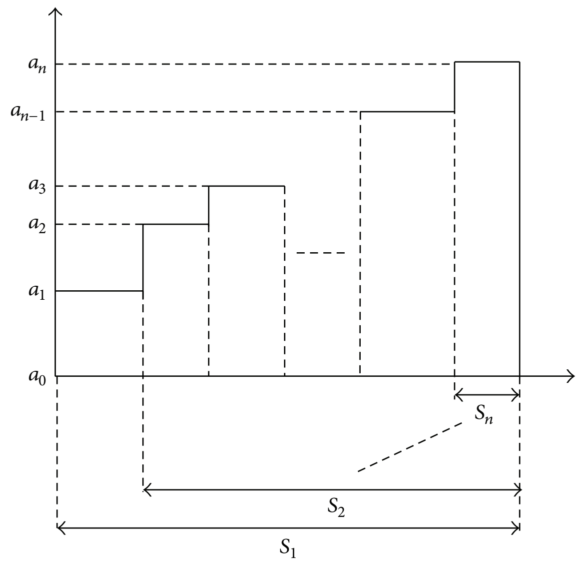


Figure 1: Choquet integral.

Definition 13 (symmetric Choquet integral [20]). Given a fuzzy measure $\mu : 2^N \to [0,1]$ and a function f on N, the symmetric Choquet integral of f with respect to μ is defined by

$$\begin{aligned}
(\mathrm{SC})\int f\,d\mu := &-\int_0^{+\infty}\mu(\{x \mid f^-(x) > r\})\,dr\\
&+\int_0^{+\infty}\mu(\{x \mid f^+(x) > r\})\,dr,
\end{aligned} \tag{10}$$

where $f^+ := f \vee 0$, $f^- := -(f \wedge 0)$.

Proposition 14. *Consider*

$$\begin{aligned}
&(\mathrm{ASC})\int f\,d\mu\\
&\quad = -\int_{-\infty}^{0}\mu^d(\{x \mid f^-(x) > r\})\,dr\\
&\qquad +\int_0^{+\infty}\mu(\{x \mid f^+(x) > r\})\,dr.
\end{aligned} \tag{11}$$

Proof. Because

$$\begin{aligned}
&\int_{-\infty}^{0} (\mu(\{x \mid f(x) > r\}) - \mu(N))\, dr \\
&= \int (\mu(\{x \mid -f^{-}(x) > r\}) - \mu(N))\, dr \\
&= -\int (\mu(\{x \mid f^{-}(x) \le r\}) - \mu(N))\, dr \\
&= -\int \mu^{d}(\{x \mid f^{-}(x) > r\})\, dr.
\end{aligned} \tag{12}$$

□

by this proposition, we can say that the difference between the asymmetric Choquet integral and the symmetric Choquet integral is based on which measure are used for integrating the negative part of f.

3. Description of Fairness by the Choquet Integral

In this section, we discuss the generalized Choquet integrals in terms of describing fairness by the Choquet integral. Let μ be a fuzzy measure on 2^N and $f = (x_1, x_2, \dots, x_n)$ a function on N. For example, N corresponds to a set consisting of n players and f corresponds to a distribution of resources, where f can take negative values. The more players are satisfied, the fairer the distribution is.

Example 15. Let $n = 3$ and there are 6 resources. Then, $f_1 = (2, 2, 2)$ is fairer than $f_2 = (4, 2, 0)$ and f_2 is fairer than $f_3 = (6, 0, 0)$. To describe this, we use the Choquet integral with respect to a submodular function. Suppose that $\mu(S)$ takes the same values according to $|S|$ (note that in this case the Choquet integral becomes an OWA operator and properties like submodularity depend on the choice of OWA weights). If μ is submodular, then we have

$$(\mathrm{C}) \int f_1 d\mu \ge (\mathrm{C}) \int f_2 d\mu \ge (\mathrm{C}) \int f_3 d\mu. \tag{13}$$

Next, we consider the negative distribution.

Example 16. Let $n = 3$ and there are -6 resources. Then, $f_1 = (-2, -2, -2)$ is fairer than $f_2 = (-4, -2, 0)$ and f_2 is fairer than $f_3 = (-6, 0, 0)$. Let μ be a submodular function. If we use the symmetric Choquet integral, then we have

$$(\mathrm{SC}) \int f_1 d\mu \le (\mathrm{SC}) \int f_2 d\mu \le (\mathrm{SC}) \int f_3 d\mu, \tag{14}$$

where all integrals take negative values. And if μ' would be another but superadditive fuzzy measure, then we obtain

$$(\mathrm{SC}) \int f_1 d\mu' \ge (\mathrm{SC}) \int f_2 d\mu' \ge (\mathrm{SC}) \int f_3 d\mu' \tag{15}$$

which is not what we wanted. In other words, we can never describe both positive fairness and negative fairness by one fuzzy measure using the symmetric Choquet integral. On the other hand, using the asymmetric Choquet integral with respect to μ where μ^d is superadditive, we obtain

$$(\mathrm{ASC}) \int f_1 d\mu \ge (\mathrm{ASC}) \int f_2 d\mu \ge (\mathrm{ASC}) \int f_3 d\mu. \tag{16}$$

If μ is a supermodular fuzzy measure, then μ is superadditive and μ^d is subadditive. Therefore, using the asymmetric Choquet integral with respect to a supermodular fuzzy measure, we can describe both positive fairness and negative fairness by one fuzzy measure.

According to these examples, the asymmetric Choquet integral with respect to a superadditive fuzzy measure is better suited for describing fairness. The Möbius transform enables us to construct fuzzy measure easily.

Definition 17 (Möbius transform). The Möbius transform of μ, denoted by $m^{\mu} : 2^N \to [-1, 1]$, is defined by

$$m^{\mu}(S) := \sum_{T \subseteq S} (-1)^{|S \setminus T|} \mu(T) \tag{17}$$

for any $S \in 2^N$. Conversely, we obtain

$$\mu(S) = \sum_{T \subseteq S} m^{\mu}(T) \tag{18}$$

by the inverse Möbius transform for any $S \in 2^N$, and there is a one-to-one correspondence between μ and m^{μ}.

Note that if the sum of all Möbius masses is 1, then the corresponding fuzzy measure of the whole set is 1.

Proposition 18. *If $m^{\mu}(S) \ge 0$ for any $S \in 2^N$, then μ is supermodular.*

Proof. Assume, for any $S \in 2^N$, $m^{\mu}(S) \ge 0$. Then, denoting that $r := S \setminus T, s := S \cap T, t := T \setminus S$ for $S, T \in 2^N$, we have

$$\begin{aligned}
&\mu(S \cup T) + \mu(S \cap T) \\
&= \sum_{C \subseteq S \cup T} m^{\mu}(U) + \sum_{U \subseteq S \cap T} m^{\mu}(U) \\
&= \Bigg(\sum_{U \subseteq r} + \sum_{U \subseteq s} + \sum_{U \subseteq t} + \sum_{\substack{U \subseteq r \cup s, U \cap r \neq \emptyset, \\ U \cap s \neq \emptyset}}
\end{aligned}$$

$$+\sum_{\substack{U\subseteq s\cup t, U\cap s\neq\emptyset,\\ U\cap t\neq\emptyset}}+\sum_{\substack{U\subseteq r\cup t,\\ U\cap r\neq\emptyset, U\cap t\neq\emptyset}}$$

$$\left.+\sum_{\substack{U\subseteq r\cup s\cup t, U\cap r\neq\emptyset,\\ U\cap s\neq\emptyset, U\cap t\neq\emptyset}}\right)m^{\mu}(U)+\sum_{U\subseteq s}m^{\mu}(U),$$

$$\mu(S)+\mu(T)=\sum_{U\subseteq S}m^{\mu}(U)+\sum_{U\subseteq T}m^{\mu}(U)$$

$$=\left(\sum_{U\subseteq r}+\sum_{U\subseteq s}+\sum_{\substack{U\subseteq r\cup s, U\cap r\neq\emptyset,\\ U\cap s\neq\emptyset}}\right)m^{\mu}(U)$$

$$+\left(\sum_{U\subseteq s}+\sum_{U\subseteq t}+\sum_{\substack{U\subseteq s\cup t, U\cap s\neq\emptyset,\\ U\cap t\neq\emptyset}}\right)m^{\mu}(U), \tag{19}$$

so that we obtain

$$(\mu(S\cup T)+\mu(s))-(\mu(S)+\mu(T))$$

$$=\left(\sum_{\substack{U\subseteq r\cup t,\\ U\cap r\neq\emptyset, U\cap t\neq\emptyset}}\right. \tag{20}$$

$$\left.+\sum_{\substack{U\subseteq r\cup s\cup t, U\cap r\neq\emptyset,\\ U\cap s\neq\emptyset, U\cap t\neq\emptyset}}\right)m^{\mu}(U)\geq 0.$$

□

Remark 19. The converse of Proposition 18 is not necessarily true. In fact, define m^{μ} on $N=\{1,2,3\}$ by

$$m^{\mu}(S):=\begin{cases}0, & S=\emptyset,\\ 1, & S\neq\emptyset,\{1,2,3\},\\ -\dfrac{1}{2}, & S=\{1,2,3\}.\end{cases} \tag{21}$$

Then, we obtain a supermodular fuzzy measure,

$$\mu(\emptyset)=0,\qquad \mu(\{1\})=\mu(\{2\})=\mu(\{3\})=1,$$
$$\mu(\{1,2\})=\mu(\{1,3\})=\mu(\{2,3\})=3, \tag{22}$$
$$\mu(\{1,2,3\})=\frac{11}{2}.$$

4. Fairness Relation Using the Asymmetric Choquet Integral

Definition 20 (Choquet Integral relation). Let $x=(x_1, x_2,\ldots,x_n)$, let $y=(y_1,y_2,\ldots,y_n)$, and let μ be a fuzzy measure. x is said to CI-dominate y, denoted by $x\geq_{\mu}y$, if and only if

$$(\mathrm{ASC})\int(x-y)\,d\mu\geq\theta. \tag{23}$$

Assume N is a set of all players x and y are distributions of resources to all players. Then, $x-y$ means each degree of satisfaction of the distribution x compared with the distribution y. Since the more players are satisfied, the fairer the distribution is, $x\geq_{\mu}y$ means x is at least as fair as y. By $x>_{\mu}y$ we denote the corresponding strict relation.

Proposition 21. *Let μ be a superadditive fuzzy measure on 2^N. Then, it is an antisymmetric relation; that is, $x\geq_{\mu}y$ and $y\geq_{\mu}x$ implies $x=y$ for any functions x, y on N.*

Proposition 22. *If μ is a supermodular measure on 2^N, then $>_{\mu}$ is a transitive relation; that is, if $x\geq_{\mu}y$ and $y\geq_{\mu}z$, then $x\geq_{\mu}z$ for any functions x, y, z on N.*

By Propositions 21 and 22, using the asymmetric Choquet integral with respect to supermodular functions, we obtain a fair ordering relation.

The proofs for Propositions 21 and 22 will be given in the Appendix of this paper.

5. Decision Making by CI-Fairness: A Numerical Example

In this section, we want to demonstrate how CI-fairness relation can be used for decision making. Consider a situation where resources have to be allocated to 3 agents A, B, and C, but sets of agents can utilize resources differently (e.g., giving a commodity to one agent, how many items can she produce within a time unit?). Such resource utilization can be assessed by a fuzzy measure:

subset S	$\mu(S)$
$\emptyset$	0
A	0.1
B	0.3
C	0.5
A, B	0.5
A, C	0.5
B, C	0.8
A, B, C	1

(24)

Here, for example, C appears to be the best in resource utilization, but A is good in collaboration. The measure is superadditive. We can also see that this is a supermodular measure. There are 3 cases to check; the other follow from subsethood or superadditivity:

(1) $\mu(\{A,B\}\cup\{A,C\})+\mu(\{A,B\}\cap\{A,C\})=\mu(\{A,B,C\})+\mu(\{A\})=1.1\geq\mu(\{A,B\})+\mu(\{A,C\})=1.0$;

(2) $\mu(\{A,B\}\cup\{B,C\})+\mu(\{A,B\}\cap\{B,C\})=\mu(\{A,B,C\})+\mu(\{B\})=1.3\geq\mu(\{A,B\})+\mu(\{B,C\})=1.3$;

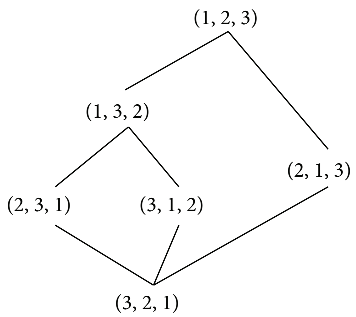


FIGURE 2: CI-Fairness relation among six possible allocations. The connections include transitivity.

(3) $\mu(\{A, C\} \cup \{B, C\}) + \mu(\{A, C\} \cap \{B, C\}) = \mu(\{A, B, C\}) + \mu(\{C\}) = 1.5 \geq \mu(\{A, C\}) + \mu(\{B, C\}) = 1.3$;

$$\begin{gathered}
(1,2,3) \geq_\mu (1,3,2) \\
(1,2,3) \geq_\mu (2,1,3) \\
(1,2,3) \geq_\mu (2,3,1) \\
(1,2,3) \geq_\mu (3,1,2) \\
(1,2,3) \geq_\mu (3,2,1) \\
(1,3,2) \geq_\mu (2,3,1) \\
(1,3,2) \geq_\mu (3,1,2) \\
(1,3,2) \geq_\mu (3,2,1) \\
(2,1,3) \geq_\mu (2,3,1) \\
(2,1,3) \geq_\mu (3,1,2) \\
(2,1,3) \geq_\mu (3,2,1) \\
(2,3,1) \geq_\mu (3,2,1) \\
(3,1,2) \geq_\mu (3,2,1) .
\end{gathered} \tag{25}$$

We find the allocation $(1, 2, 3)$ in relation to all other allocations, but no allocation is in relation to $(1, 2, 3)$. This is natural, given the order of resource utilization weights for the three agents. By rank, we can now also order the other 5 allocations.

Now, we want to allocate three shares of a resource of 6 units $(1, 2, 3)$ among A, B, C and compare the 6 possible permutations of $(1, 2, 3)$ by the Choquet integral relation with $\theta = 0$. For example, to see if allocation $x = (1, 2, 3)$ is in relation to the allocation $y = (2, 3, 1)$, we have to test

$$(C) \int [x - y]^+ d\mu \geq (C) \int [y - x]^+ d\mu^d. \tag{26}$$

In this case, $[x - y]^+ = (0, 0, 2)$ and $[y - x]^+ = (1, 1, 0)$. So, the first integral becomes $2 \times \mu(\{C\}) = 1.0$ and the second integral $1 \times \mu^d(\{A, B\}) = (1 - \mu(\{C\})) = 0.5$. From $1.0 \geq 0.5$, then $x \geq_\mu y$ follows.

Among all possible 30 pairs (x, y) with $x \neq y$, we can find 13 cases in total where the Choquet integral relation holds.

Figure 2 shows the Hasse diagram of the relation for the 6 possible allocations. We can see that, as an effect of the better collaboration weight of A, allocation $(2, 1, 3)$ is promoted and appears on the same rank as the strong allocation $(1, 3, 2)$ (strong with regard to single agent utilization of a resource) and also of better rank than the allocation $(2, 3, 1)$—expressing the preference for C as partner of A when comparing possible wins and losses in the allocation. Finally, as expected, there is no case to allocate 3 shares to A.

6. Experiments

In this section, we will show how the CI-fairness can be applied in a practical situation. We study wireless channel allocation in the network layer; a base station BS that is connected to the backbone can command over a number m of wireless transmission *channels* and allocates them to a number n of users (or subscriber stations, mobile stations, relay stations, etc.—depending on the specific network architecture) for uplink or downlink traffic. In the most abstract notation, the BS makes the allocation based on a measured or estimated *channel state information*; if it allocates a channel i to user j there will be a *channel coefficient* C_{ij} from $[0, 1]$ that represents (in a simplified form) to what degree the user can employ that channel. For example, remote users are likely to have smaller channel coefficients than users that are close to the BS. Here, an allocation is seen as *feasible* if each channel is allocated to one user and to each user at least one channel is assigned.

The sum of all channel coefficients of allocated channels for a specific user is seen as the *performance* of the allocation for that user. Then, the optimality task is related to the performance vector for all users. This is a direct reference to a relation between performance vectors, and various fairness relations can be easily and conveniently studied in this context. If the number of channels and users is not too large, exhaustive search for maximum sets is possible. For example, for 4 users and 6 channels, there are 1560 feasible allocations, which means about 2.4 Mio. pairwise comparisons.

In this model, channel allocation ignores low channel interferences between users and specific channels by the restriction that each channel is allocated to exactly one user. However, if implementing a wireless access where some users can use the same channel for transmission, it might be hard to decide the exact schedule of such multiple assignments, or to decide if installing such a scheme at all. Therefore, we follow a more simplified approach; channel interference will be represented by a conflict matrix between users, where low elements for pairs of users indicate the option to use multichannel assignment. Then, we use the converse elements of the conflict matrix as masses in the Möbius transform. More specifically, the Möbius masses for single element sets are set to equal values, the masses for two-element are set to 1 minus the corresponding element of the conflict matrix, and all other are set to 0. Then, the massvector is normalized to a total of all masses of 1 to ensure that

Table 1: Comparison of total performances for capacity, proportional, and maxmin fairness.

Fairness relation	Average performance
Conflict matrix range 0 to 1	
CI	4.2882
Proportional	4.54661
maxmin	4.14688
Conflict matrix range 0.9 to 1	
CI	4.56051
Proportional	4.51135
maxmin	4.22561
Conflict matrix range 0 to 0.2	
CI	4.09365
Proportional	4.41848
maxmin	4.04879

for the fuzzy measure the measure for the whole set is 1. The inverse Möbius transform of such a mass vector gives a supermodular measure that can be used to specify a CI-based fairness relation according to Definition 20.

Now, this relation is applied to all feasible channel allocations—it means we are not considering a specific use of multichannel assignment but we are looking for the influence that such a multichannel allocation could have on the specific selection of maximal elements. Then, the focus is on the disadvantage that some pairs of users can experience if their favoured situation (e.g., to be distant enough to use the same channel) is not taken into account. The similar approach can be considered for cognitive radio, opportunistic networking, P2P, vehicular networks, and so forth.

We might especially look for the *price of fairness* [1], compared to a standard proportional fair or maxmin fair allocation.

Some example results are shown in Table 1. There, for the case of 4 users and 6 cells, the maximum sets over all feasible allocations were computed for CI-fairness using a measure as described above, proportional fairness, and maxmin fairness. The conflict matrix elements for the CI-fairness were uniform randomly set within specific ranges. All results are averaged over 30 repetitions. Note that average sizes of maximum sets were found to be 4.2 for maxmin fairness, 6.1 for proportional fairness, and 41.9 for CI-fairness, so the CI-fairness produces larger maximum sets, an issue that should be addressed in future works. The table now shows average performances over all maximal elements.

We see with regard to efficiency the well-known fact that proportional-fair allocations are in average more efficient than maxmin fair allocations. The CI-fairness appears to select maximal relations with a total performance between proportional fairness and maxmin fairness. In case of strong interferences (where a multichannel assignment is not reasonable), it is very close to proportional fairness, while in case of low interference it appears to be more close to maxmin fairness, that is, the fairness relation favouring least elements. Thus, such an empirical result can be understood in the sense that the neglecting of the potential multichannel assignment corresponds with the neglection of strong users. This is also particularly appearing for the general case, where all conflict matrix elements were randomly selected between 0 and 1; also here, the average performance appears more close to maxmin fairness than proportional fairness.

7. Conclusion

A fairness relation among vectors based on the asymmetric Choquet integral was studied. It formally follows the Nash standard of comparison where the relative losses and wins are replaced by absolute losses and wins, but instead taking weights for groups of vector components into account. Thus, it can represent, for example, pairs of users that are in a favourable situation regarding resource allocation. The appealing points of this CI-relation are as follows. (1) In case a supermodular measure is used for the integration, the relation will be transitive, which suits faster search for maximal elements; (2) it can be parametrized by giving a weight to each subset of vector components (while in practice this might be relevant only for smaller subsets). The representation of the fuzzy measure by its Möbius transform appears to be a convenient way to yield a supermodular fuzzy measure; the masses just need to be all nonnegative. Thus, the relation can be conveniently applied in many practical applications to specify optimality. Slight disadvantages are with the complexity of the involved calculations, and some empirical evidence for specifying larger maximum sets than other fairness relations like proportional fairness and maxmin fairness.

In case of wireless channel allocation, how CI-fairness can be used to help deciding whether implementations of more complex allocation schemes are indeed worth the effort was suggested. This method can be applied to many other problems of higher efficiency of network resource utilization. In future work, we will consider the use of advanced fuzzy integrals like balancing and fusion Choquet integral and the adjustment of related CI-fairness parameters to specific situations and what can be concluded from such parameters. We will also more intensively study the performance issues with regard to computational effort and search for maximal elements and study conditions where also greatest elements exist.

Appendix

Proofs of Theorems

In the following, we provide the proofs for Propositions 21 and 22. For formal convenience of the proofs, the notation will be a little bit changing; therefore, we repeat the basic definitions.

We define the Choquet-integral-based (fairness) relation as follows: be x and y two vectors from R^n. By $[x]^+$, we indicate the positive support of a vector x; that is, the ith component of $[x]^+$ is x_i if $x_i \geq 0$ or 0 if $x_i < 0$. The Choquet integral is denoted by $(C)\int x\,d\mu$. There, μ is a fuzzy measure.

Also, μ^d denotes its dual measure (i.e., $\mu^d(A) = \mu_X - \mu(A^C)$). Then, x is said to CI-dominate y if and only if

$$x \geq_\mu y \longleftrightarrow (C) \int [x-y]^+ d\mu \geq (C) \int [y-x]^+ d\mu^d. \quad \text{(A.1)}$$

We can alternatively rewrite this as the asymmetric Choquet integral:

$$x \geq_\mu y \longleftrightarrow (\text{ASC}) \int (x-y)\, d\mu \geq 0 \quad \text{(A.2)}$$

since the right-hand side condition of (A.1) exactly corresponds with the definition of the asymmetric Choquet integral.

First, we can show the following.

Lemma 23. *If μ is a superadditive fuzzy measure where for each $A, B \neq \emptyset$ and $A \cap B = \emptyset$ $\mu_A + \mu_B < \mu_{A\cup B}$, then for $A \neq \emptyset$ $\mu_A^d > \mu_A$.*

Proof. Since μ is superadditive, for $A \neq \emptyset$ $\mu_A + \mu_{A^C} < \mu_{A\cup A^C} = \mu_X$. Then, $\mu_A < \mu_X - \mu_{A^C} = \mu_A^d$ follows directly. □

Theorem 24. *If μ is superadditive, then from $x \geq_\mu y$ and $y \geq_\mu x x = y$ follows.*

Proof. $x \geq_\mu y$ means $(C) \int [x-y]^+ d\mu \geq (C) \int [y-x]^+ d\mu^d$. If $x \neq y$, we can use Lemma 23 to obtain

$$\begin{aligned}(C) \int [x-y]^+ d\mu^d &> (C) \int [x-y]^+ d\mu \\ \geq (C) \int [y-x]^+ d\mu^d &> (C) \int [y-x]^+ d\mu;\end{aligned} \quad \text{(A.3)}$$

that is, $(C) \int [y-x]^+ d\mu \ngeq (C) \int [x-y]^+ d\mu^d$. This would mean that $y \geq_\mu x$ would not hold. Therefore, it must be $x = y$. □

For a supermodular measure and nonnegative vectors x and y, we have the known property $(C) \int x\, d\mu + (C) \int y\, d\mu \leq (C) \int (x+y) d\mu$. Since the dual measure of a supermodular measure is submodular, it also holds that $(C) \int x\, d\mu + (C) \int y\, d\mu \geq (C) \int (x+y) d\mu$. Using this, the following can be shown.

Theorem 25. *If μ is supermodular, then the relation $>_\mu$ is cycle-free.*

Proof. We will only show the case of nonexistence of a 3-cycle; here, the concept generalizes directly to any other case. Assume that $x >_\mu y$ and $y >_\mu z$. First, we note that

$$\begin{aligned}&[x-y]^+ + [y-z]^+ + [z-x]^+ \\ &\quad = [y-x]^+ + [z-y]^+ + [x-z]^+,\end{aligned} \quad \text{(A.4)}$$

and therefore

$$\begin{aligned}&(C) \int \left([x-y]^+ + [y-z]^+ + [z-x]^+\right) d\mu \\ &\quad = (C) \int \left([y-x]^+ + [z-y]^+ + [x-z]^+\right) d\mu.\end{aligned} \quad \text{(A.5)}$$

Using Lemma 23, we yield

$$\begin{aligned}&(C) \int \left([y-x]^+ + [z-y]^+ + [x-z]^+\right) d\mu^d \\ &\quad > (C) \int \left([x-y]^+ + [y-z]^+ + [z-x]^+\right) d\mu,\end{aligned} \quad \text{(A.6)}$$

and from supermodularity of μ and submodularity of μ^d

$$\begin{aligned}&(C) \int [y-x]^+ d\mu^d + (C) \int [z-y]^+ d\mu^d + (C) \int [x-z]^+ d\mu^d \\ &\quad > (C) \int [x-y]^+ d\mu + (C) \int [y-z]^+ d\mu \\ &\qquad + (C) \int [z-x]^+ d\mu.\end{aligned} \quad \text{(A.7)}$$

However, $x >_\mu y$ means $(C) \int [x-y]^+ d\mu \geq (C) \int [y-x]^+ d\mu^d$ and $y >_\mu z$ means $(C) \int [y-z]^+ d\mu \geq (C) \int [z-y]^+ d\mu^d$. Therefore, above inequality can only hold if

$$(C) \int [z-x]^+ d\mu < (C) \int [x-z]^+ d\mu^d \quad \text{(A.8)}$$

which means $z \not>_\mu x$. □

If the measure μ is supermodular, also transitivity of the relation can be shown. Before we can show this, we need to introduce a notation and a corresponding lemma. The notation refers to computing the (asymmetric) Choquet integral of x by measure μ in a "different order." If $x_{x(i)}$ indicates the ith largest element of x, then the Choquet integral is calculated as

$$\begin{aligned}(C) \int x d\mu = {}& x_{x(1)} \mu_{x(1)} \\ &+ \sum_{i=2}^{n} x_{x(i)} \left[\mu_{x(1)x(2)\cdots x(i)} - \mu_{x(1)\cdots x(i-1)}\right].\end{aligned} \quad \text{(A.9)}$$

For any other y, we can compute the same expression, just following the order of the elements of y instead of x. Then, $x_{y(i)}$ indicates the element of x with the index of the ith largest element of y, and we define (for convenience, we still use the symbol for the Choquet integral, keeping in mind that it is not a real fuzzy integral anymore):

$$\begin{aligned}(C) \int_{(y)} x\, d\mu = {}& x_{y(1)} \mu_{y(1)} \\ &+ \sum_{i=2}^{n} x_{y(i)} \left[\mu_{y(1)y(2)\cdots y(i)} - \mu_{y(1)\cdots y(i-1)}\right].\end{aligned} \quad \text{(A.10)}$$

Consequently, we could also write $(C)\int_{(x)} x d\mu$ for the "original" Choquet integral. We can do the same for the asymmetric Choquet integral, by applying the resorting to the negative components of the vector x as well, and will use the notation $(\text{ASC})\int_{(y)} x\, d\mu$ here. Now, we have the following.

Lemma 26. *If μ is a supermodular measure, then for any y $(\text{ASC})\int_{(y)} x\, d\mu \geq (\text{ASC})\int_{(x)} x\, d\mu$.*

Proof. A moment of reasoning gives that any sorting of the indizes can be achieved by suitable application of a sequence of three kinds of "neighbour swaps", starting from the order $x(1)x(2)\cdots x(n)$).

Swap 1. In the computation of the integral, a larger or equal nonnegative element d_i at index i is swapped with its immediate nonnegative neighbour d_{i+1} to the right. Then, what was computed as $d_i(\mu_{a,i}-\mu_a)+d_{i+1}(\mu_{a,i,i+1}-\mu_{a,i})$ before the swap changes to $d_{i+1}(\mu_{a,i+1}-\mu_a)+d_i(\mu_{a,i,i+1}-\mu_{a,i+1})$ after the swap, all other parts of the resorted asymmetric Choquet integral expression remain the same (by a we indicate the index order before index i). Thus, the total change is

$$\begin{aligned} & d_i\left(\mu_{a,i,i+1}-\mu_{a,i+1}-\mu_{a,i}+\mu_{a,i}\right) \\ & \quad - d_{i+1}\left(\mu_{a,i,i+1}-\mu_{a,i}-\mu_{a,i+1}+\mu_{a,i}\right), \end{aligned} \tag{A.11}$$

and from $d_i \geq d_{i+1}$ and $\mu_{a,i,i+1}+\mu_a \geq \mu_{a,i+1}+\mu_{a,i}$ (supermodularity of μ) it directly follows that the computed value will not decrease by this swap.

Swap 2. In the computation of the integral, a larger or equal negative element d_{i+1} at index $i+1$ is swapped with its immediate negative left neighbour d_i. From a similar evaluation as for *Swap 1*, it can be seen that also here, the value of the computed expression will not decrease after the swap.

Swap 3. A nonnegative element at index i is swapped with its immediate negative neighbor to the right with index $(i+1)$. Then, before the swap, we compute $d_i(\mu_{a,i}-\mu_a) - d_{i+1}(\mu_{a,i,i+1}-\mu_{a,i})$ (d_{i+1} is the absolute value of the negative-valued neighbour of d_i) and after the swap this part of the expression changes to $-d_{i+1}(\mu_{a,i+1}-\mu_a)+d_i(\mu_{a,i,i+1}-\mu_{a,i+1})$. Then, the total change is

$$\begin{aligned} & d_i\left(\mu_{a,i,i+1}-\mu_{a,i+1}-\mu_{a,i}+\mu_a\right) \\ & \quad - d_{i+1}\left(\mu_{a,i+1}-\mu_a-\mu_{a,i,i+1}+\mu_{a,i}\right). \end{aligned} \tag{A.12}$$

Supermodularity of μ gives $\mu_{a,i}+\mu_{a,i+1} \leq \mu_a+\mu_{a,i,i+1}$ and therefore the factor of d_i in former expression is nonnegative and the factor of d_{i+1} (which is the negation of the factor of d_i) is negative or 0. Therefore, the total change is nonnegative as well.

Thus, with each of these swap operations, we will never decrease the value of the expression, and starting from the asymmetric Choquet integral for x upon reaching the final order the final value will be larger or equal. □

An example might be helpful. Assume that the order in x is $(1\ 2\ 3\ -4\ -5)$ (a negative index should indicate that the corresponding element of x is negative) and in y we have the order $(-5\ 3\ 1\ 2\ -4)$ (so, y_5 is the largest component of y and y_4 the smallest). Then, the order of swaps is as follows: *Swap* 1 applied as $(1\ 2\ 3\ -4\ -5)$ gives $(1\ 3\ 2\ -4\ -5)$; *Swap* 1 applied as $(1\ 3\ 2\ -4\ -5)$ gives $(3\ 1\ 2\ -4\ -5)$. Next, *Swap* 2 applied as $(3\ 1\ 2\ -4\ -5)$ gives $(3\ 1\ 2\ -5\ -4)$, *Swap* 3 applied as $(3\ 1\ 2\ -5\ -4)$ gives $(3\ 1\ -5\ 2\ -4)$, *Swap* 3 applied as $(3\ 1\ -5\ 2\ -4)$ gives $(3\ -5\ 1\ 2\ -4)$, and finally *Swap* 3 applied as $(3\ -5\ 1\ 2\ -4)$ gives the order $(-5\ 3\ 1\ 2\ -4)$ of y. Note also that in case the elements of x are not in that order we can always relabel the indizes and measures correspondingly.

This means if we compute (using $x = (x_1, x_2, x_3, -x_4, -x_5)$)

$$\begin{aligned} (\text{ASC}) & \int_{(y)} x\, d\mu \\ & = -x_5\mu_5 + x_3\left(\mu_{3,5}-\mu_5\right) + x_1\left(\mu_{1,3,5}-\mu_{3,5}\right) \\ & \quad + x_2\left(\mu_{1,2,3,5}-\mu_{1,3,5}\right) - x_4\left(\mu_{1,2,3,4,5}-\mu_{1,2,3,5}\right) \end{aligned} \tag{A.13}$$

instead of

$$\begin{aligned} (\text{ASC}) & \int x\, d\mu \\ & = x_1\mu_1 + x_2\left(\mu_{1,2}-\mu_1\right) + x_3\left(\mu_{1,2,3}-\mu_{1,2}\right) \\ & \quad - x_4\left(\mu_{1,2,3,4}-\mu_{1,2,3}\right) - x_5\left(\mu_{1,2,3,4,5}-\mu_{1,2,3,4}\right), \end{aligned} \tag{A.14}$$

then $(\text{ASC})\int_{(y)} x\, d\mu \geq (\text{ASC})\int x\, d\mu$.

Using this Lemma, we can now easily show the following.

Theorem 27. *If μ is a supermodular measure, then $\geq_\mu$ is a transitive relation.*

Proof. For transitivity, we have to show that for any x, y, and z such that $x \geq_\mu y$ and $y \geq_\mu z$ it follows that $x \geq_\mu z$. Now, $x \geq_\mu y$ means $(\text{ASC})\int (x-y) d\mu \geq 0$ and $y \geq_\mu z$ means $(\text{ASC})\int (y-z) d\mu \geq 0$. From supermodularity of μ and Lemma 26, then it also follows that

$$\begin{aligned} (\text{ASC})\int_{(x-z)} (x-y)\, d\mu \geq (\text{ASC})\int (x-y)\, d\mu \geq 0, \\ (\text{ASC})\int_{(x-z)} (y-z)\, d\mu \geq (\text{ASC})\int (y-z)\, d\mu \geq 0. \end{aligned} \tag{A.15}$$

However, by using the fact that

$$\begin{aligned} (\text{ASC}) & \int_{(x-z)} (x-y)\, d\mu + (\text{ASC})\int_{(x-z)} (y-z)\, d\mu \\ & = (\text{ASC})\int (x-z)\, d\mu, \end{aligned} \tag{A.16}$$

we see that also

$$(\text{ASC})\int (x-z)\, d\mu \geq 0 \tag{A.17}$$

which means $x \geq_\mu z$. □

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," *Operations Research*, vol. 59, no. 1, pp. 17–31, 2011.

[2] J. F. Nash Jr., "The bargaining problem," *Econometrica*, vol. 18, pp. 155–162, 1950.

[3] E. Kalai and M. Smorodinsky, "Other solutions to Nash's bargaining problem," *Econometrica*, vol. 43, pp. 513–518, 1975.

[4] J. Rawls, *A Theory of Justice*, Harvard University Press, 1999.

[5] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.

[6] D. P. Bertsekas, R. G. Gallager, and P. Humblet, *Data Networks*, vol. 2, Prentice-Hall International, 1992.

[7] H. P. Young, *Equity: In Theory and Practice*, Princeton University Press, Princeton, NJ, USA, 1995.

[8] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.

[9] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proceedings of the 11th Annual International Conference on Mobile Computing and Networking (MobiCom '05)*, pp. 58–72, ACM, Cologne, Germany, August-September 2005.

[10] A. Iyer, C. Rosenberg, and A. Karnik, "What is the right model for wireless channel interference?" *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2662–2671, 2009.

[11] A. Mishra, V. Shrivastava, S. Banerjee, and W. Arbaugh, "Partially overlapped channels not considered harmful," *ACM SIGMETRICS Performance Evaluation Review*, vol. 34, no. 1, pp. 63–74, 2006.

[12] N. Nie and C. Comaniciu, "Adaptive channel allocation spectrum etiquette for cognitive radio networks," *Mobile Networks and Applications*, vol. 11, no. 6, pp. 779–797, 2006.

[13] C. Peng, H. Zheng, and B. Y. Zhao, "Utilization and fairness in spectrum assignment for opportunistic spectrum access," *Mobile Networks and Applications*, vol. 11, no. 4, pp. 555–576, 2006.

[14] X. Yang and A. P. Petropulu, "Co-channel interference modeling and analysis in a Poisson field of interferers in wireless communications," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 64–76, 2003.

[15] M. Köppen, "Relational optimization and its application: from bottleneck flow control to wireless channel allocation," *Informatica*, vol. 24, no. 3, pp. 413–433, 2013.

[16] M. Sugeno, *Theory of fuzzy integrals and its applications [Ph.D. thesis]*, 1974.

[17] S. Michio, "Fuzzy measures and fuzzy integrals: a survey," *Fuzzy Automata and Decision Processes*, vol. 78, no. 33, pp. 89–102, 1977.

[18] D. Denneberg, *Non-Additive Measure and Integral*, vol. 27, Springer, 1994.

[19] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1954.

[20] J. Šipoš, "Integral with respect to a pre-measure," *Mathematica Slovaca*, vol. 29, no. 2, pp. 141–155, 1979.

[21] M. Grabisch and C. Labreuche, "The symmetric and asymmetric Choquet integrals on finite spaces for decision making," *Statistical Papers*, vol. 43, no. 1, pp. 37–52, 2002.

[22] M. Grabisch, C. Labreuche, and J.-C. Vansnick, "On the extension of pseudo-Boolean functions for the aggregation of interacting criteria," *European Journal of Operational Research*, vol. 148, no. 1, pp. 28–47, 2003.

[23] R. Mesiar, A. Mesiarová-Zemánková, and K. Ahmad, "Discrete Choquet integral and some of its symmetric extensions," *Fuzzy Sets and Systems*, vol. 184, pp. 148–155, 2011.

[24] A. Mesiarová-Zemánková, R. Mesiar, and K. Ahmad, "The balancing Choquet integral," *Fuzzy Sets and Systems*, vol. 161, no. 17, pp. 2243–2255, 2010.

[25] H. Fouchal, X. Gandibleux, and F. Lehuédé, "A lower bound of the Choquet integral integrated within Martins'algorithm," in *New State of MCDM in the 21st Century*, vol. 648 of *Lecture Notes in Economics and Mathematical Systems*, pp. 79–89, Springer, Heidelberg, Germany, 2011.

[26] L. Galand, L. Julien, and P. Patrice, "Dominance rules for the choquet integral in multiobjective dynamic programming," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pp. 538–544, AAAI Press, 2013.

[27] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, Springer, 2nd edition, 2010.

[28] W. Ogryczak, A. Wierzbicki, and M. Milewski, "A multi-criteria approach to fair and efficient bandwidth allocation," *Omega*, vol. 36, no. 3, pp. 451–463, 2008.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 860747, 7 pages
http://dx.doi.org/10.1155/2014/860747

Research Article

Price of Fairness on Networked Auctions

Mariusz Kaleta

Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland

Correspondence should be addressed to Mariusz Kaleta; mkaleta@ia.pw.edu.pl

Received 8 March 2014; Revised 15 May 2014; Accepted 4 June 2014; Published 17 July 2014

Academic Editor: Hanan Luss

We consider an auction design problem under network flow constraints. We focus on pricing mechanisms that provide fair solutions, where fairness is defined in absolute and relative terms. The absolute fairness is equivalent to "no individual losses" assumption. The relative fairness can be verbalized as follows: no agent can be treated worse than any other in similar circumstances. Ensuring the fairness conditions makes only part of the social welfare available in the auction to be distributed on pure market rules. The rest of welfare must be distributed without market rules and constitutes the so-called *price of fairness*. We prove that there exists the minimum of *price of fairness* and that it is achieved when uniform unconstrained market price is used as the base price. The *price of fairness* takes into account costs of forced offers and compensations for lost profits. The final payments can be different than locational marginal pricing. That means that the widely applied locational marginal pricing mechanism does not in general minimize the *price of fairness*.

1. Introduction

Classical auction (through the whole paper by "auction" we mean closed double sealed exchange-like mechanism; in other words an "auction" is a set of trading rules for an exchange) mechanisms are based on the supply/demand curves intersection which sets accepted and rejected offers and determines the uniform market price. However, in many real-world infrastructure economies, a commodity flow is limited by the resources of some network system. This leads to a concept of the networked auctions [1]. Some examples come from electricity [2], gas [3, 4], water [5], telecommunication [6], and transport markets. Determination of auction winners not only must be based on economic grounds, but also must be aligned with network system resources, for example, transmission grid, water transmission network, telecommunication network, and road network resources.

In [1] the network winner determination problem (NWDP), an extension of classical winner determination problem (WDP) with consideration of network resources, has been introduced. This model solves only the problem of social welfare maximization, but there is still the question of how social welfare should be distributed between the market participants. In classical auctions, the marginal pricing principle is usually applied, which provides the solution acknowledged to be fair. In case of networked auctions the fairness conditions can be disrupted and uniform market pricing cannot be used.

To address fairness in the electricity markets, the locational marginal pricing (LMP) was introduced by Schweppe [7] and further developed by Hogan [8]. LMP sets the marginal prices at each node. Although this approach is widely recognized in a context of electricity markets, the LMP and similar marginal nodal pricing policies are also addressed for other markets [3, 5, 9]. The LMP has several flaws [10]. One of the main shortcomings is only partial distribution of social welfare between market participants. This means that there is some price that is paid to restore the fairness conditions.

In this paper we introduce the *price of fairness* (PoF) for networked auctions. We focus on multicommodity sealed-bids auctions. Our main theorem proves that the minimal PoF can be achieved if the unconstrained market price is used for settlements and some additional nodal components are paid at each node. This means that widely used locational marginal pricing approach does not minimize the PoF. Till now, there is very little literature devoted to fair networked auctions, and business is focused on LMP-like mechanisms. We think that our main result is to show that there is a wide interesting spectrum of fair auction mechanisms for networked systems. This spectrum should be further investigated including also aspects other than the PoF.

The paper is organised as follows. A discussion of related works is provided in Section 2. In Section 3, we introduce basic notions: the general model of networked auction with its components (a special case of winner determination problem and value mechanism) and the problem of finding the best value mechanism. The concept of fairness in the context of networked auction is introduced in Section 4. The main results of the paper are presented in Section 5. We define the *price of fairness* and analyze it in the space of possible value mechanisms. We conclude and present some interesting future directions in Section 6.

2. Related Literature

The concept of fairness plays an important role in resource allocation problems [11–13]. Although fairness is a subjective notion, it implies equity and impartiality [14]. To achieve these conditions several ways of specifying the fairness are considered in the literature. Usually, some measure of inequality is to be minimized [15, 16]. The simplest measures are based on the absolute measurement of the spread of outcomes, for example, mean or maximum absolute difference, or measurement relative to the mean outcome, for example, mean absolute deviation or Gini coefficients [17, 18]. Unfortunately, direct minimization of typical inequality measures is in contradiction with optimization of individual outcomes [19]. One can use an aggregation function to solve the problem; however this function should satisfy several requirements [20]. Various solution concepts can be achieved when logarithmic function is used to obtain fair aggregation [20]. The concept of equitably efficient solution is another approach, which was formalized in [21].

Unfortunately, all of these concepts of fairness and their underlying axiomatisation cannot be applied directly when the allocation is performed via auction mechanism. In this case, the fairness concept must take into consideration the prices of bids and thus different utility of bidders. Moreover, the incompleteness of information possessed by each bidder is also important.

Auction has been widely considered as a mechanism of resource allocation in selfish, multiagents environments [22]. Although the theoretical literature on auctions is rich and multidimensional, fairness issues are relatively rare. In [23], two notions of fairness are introduced for combinatorial auctions: the basic fairness and the extended fairness. The basic fairness is related to conditions that must be satisfied to leave each agent with the feeling that he has a fair share. The extended fairness ensures envy-free allocation, which means that each agent is at least as happy with his share as he would be with share of any other agent.

Murillo et al. [24] consider a recurrent auction model equipped with reservation price. In this setting, no selling offer below its reservation price can be winning. The incorporation of the reservation price forces the introduction of fairness considerations. The problem considered in [24] is a policy of setting the reservation prices to assure fairness condition understood as a similar probability of winning for every bidder.

Wu et al. [25] consider fairness in the sealed-bid auctions rather more on procedural level than the market clearing rules. In order to mitigate the possibilities for collusion or cheating, they consider when and how the sealed bids should be opened.

The concept of fairness for multicommodity auctions has been formalized by Toczyłowski in [26]. We follow this concept in our paper and we discuss the details in Section 4.

The term *price of fairness* has been introduced in [27]. In this work, the PoF has been analyzed in the context of general resource allocation problem. The authors define PoF as a relative system efficiency loss under fair conditions of the allocation. A characterization of the PoF for a broad spectrum of allocation problems is also provided in [27].

There are also some studies, that refer to the PoF indirectly; They consider loss of efficiency due to fair conditions, mainly in allocation problems. In most works, the price of fairness or similar notion is defined as the performance loss incurred relative to utility, in making allocations under one of several possible fairness criteria formulations. In [28] some numerical computations for efficiency loss in several network configurations are presented under consideration of proportional and max-min fairness. Butler and Williams have proved that, for certain class of facility location problem, the price of fairness is zero [29]. Mo and Walrand have studied some family of parameterised objective of bandwidth allocations [30]. The max-min fairness and proportional fairness are special cases of this family.

In [26] the fairness is defined by a set of requirements that must be satisfied in a fair solution of auction with uniform pricing. In our paper we follow this work and extend the analysis to the case of nodal pricing. Differentiation of buy and sell market prices is discussed in [31] with an application to electricity market. We derive the concept of separated prices for buying and selling from this work. However, we consider more general class of auction models.

The *price of anarchy* is somehow similar concept to the PoF [32, 33]. It measures the system efficiency loss due to selfish behaviour of the agents. The PoA and PoF are only partially overlapped. Even if the agents would play truly, the PoF can be positive. However, the increase of PoA can be related to higher PoF that results from deficiency of considered auction mechanism.

3. Networked Auctions

We consider an organized market in which the sellers and buyers submit their offers. Then, the auction mechanism is run to find the winning offers and to set the value flows. The auction rules can be divided into two steps: the winners determination and the pricing.

3.1. Winners Determination Problem. Let us assume that an infrastructure network is modeled by graph G, where V is a set of vertices and E is a set of edges [4]. We assume that graph G is defined by an incidence matrix $a = [a_{ve}]$. To make the notation simpler, we also assume that G is a connected graph. Any solution to the winners determination problem is equivalent to some flows in the G and must satisfy Kirchoff's laws. Each sell offer is related to a commodity that can be injected by a seller into the network at a given node. Similarly, any buy offer is related to a commodity that can be taken by a buyer from the network at a given node.

Definition 1 (vertex-oriented network winner determination problem, VWDP). The sellers submit the set of offers $j \in \mathcal{J} = \{1, 2, \ldots, J\}$. An offer j is a tuple $\langle c_j, q_j^{\max} \rangle$, where $c_j \geq 0$ is an offered unit price, and $q_j^{\max} \geq 0$ is the maximal offered volume of a commodity. The buyers submit a set of offers $m \in \mathcal{B} = \{1, 2, \ldots, B\}$. An offer m is a tuple $\langle e_m, d_m^{\max} \rangle$, where $e_m \geq 0$ is an offered unit price, and $d_m^{\max}$ is the maximal demanded volume of a commodity. Without loss of generality, to simplify the notation, we assume that each seller and each buyer are located in different node. Thus, the indexes of sellers and buyers can be also used for indexing the nodes. Let $V^J = \mathcal{J} \subseteq V$ be a set of seller nodes and let $V^B = \mathcal{B} \subseteq V$ be a set of buyer nodes. The vertex-oriented network winner determination problem (VWDP) is to find a set of winning offers and their volumes (q_j) and (d_m), $0 \leq q_j \leq q_j^{\max}$ and $0 \leq d_m \leq d_m^{\max}$, which is balanced $(\sum_{j \in \mathcal{J}} q_j = \sum_{m \in \mathcal{B}} d_m)$ and social welfare-maximizing under the network flow constraints:

$$\sum_{e \in E} a_{ve} f_e = \begin{cases} q_v & v \in V^J, \\ 0 & v \notin V^J \cup V^B, \\ -d_m & v \in V^B, \end{cases} \qquad (1)$$
$$\forall v \in V,$$

where f_e is the commodity flow over edge e and q_j and d_m are the accepted volumes of sell offer j and buy offer m, respectively. The VWDP maximizes the natural *utilitarian* criterion—the sum of the surpluses of all individual players. Full formulation of VWDP is a simple extension of classical auction models [26] and can be expressed as linear programme.

3.2. Value Mechanism. The maximal total social welfare $Q = \sum_{m \in M} e_m d_m - \sum_{j \in J} c_j q_j$ is achieved by solving VWDP. The buyers and sellers are selfish and tend to maximize their individual profits. Thus, we should notice that the submitted offers could be different than the market participants' true valuations, and then Q would not be a true social welfare, but rather "declared" market surplus [31].

The value mechanism is responsible for the surplus distribution. The distribution is usually expressed with the use of payment information $\mathcal{I}$, which may include market price, payments for individual offers, and other information. For instance, there are marginal prices at each node in $\mathcal{I}$ in case of LMP mechanism. The mechanism $\mathcal{M}$ is defined as the following mapping:

$$\mathcal{M} : (e, c, d, q) \longrightarrow \mathcal{I}. \qquad (2)$$

A very generic approach to modeling the value mechanisms was formulated by Toczyłowski and called *parametric balancing model* [26]. It introduces a virtual node denoted by 0 and connected to each other node. There are market price π_0 at virtual node and nodal prices π_v at every node in V. If commodity is injected to the network at node $k \in V$ and π_k is the nodal price, then $\pi_{k0} = \pi_0 - \pi_k$ is the unit cost of commodity transport from node k to the virtual node 0. Similarly, if node $l \in V$ is a consumption node with nodal price π_l then $\pi_{0l} = \pi_l - \pi_0$ is the unit cost for commodity transport from the virtual node 0 to node l (see Figure 1).

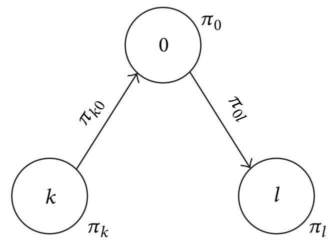


FIGURE 1: Model of pricing with nodal prices π_k, π_l, virtual node price π_0, and transmission prices π_{k0} and π_{0l}.

3.3. Problem of Finding the Best Value Mechanism. Let us assume that there is a set of L quality measures $\{Q_1(\mathcal{M}), \ldots, Q_L(\mathcal{M})\}$ of pricing mechanisms, where function $Q_l : \mathcal{M} \rightarrow \mathbb{R}$ is such that if $Q_l(\mathcal{M}_1) > Q_l(\mathcal{M}_2)$, then mechanism $\mathcal{M}_1$ is strictly preferred over the mechanism $\mathcal{M}_2$ according to the measure l. Quality measures introduce the partial order to the space of mechanisms. Then, the problem of finding the best mechanism is defined as follows:

$$\max_{\mathcal{M}} \{Q_1(\mathcal{M}), \ldots, Q_L(\mathcal{M})\}. \qquad (3)$$

For a given pricing mechanism it is interesting whether it is nondominated solution of multicriteria problem (3). In the next sections we will formulate a bit narrower space of possible mechanisms and we will analyze possible pricing mechanisms according to the PoF criterion.

4. Concept of Fairness

For a given set of nodal prices (π_v) the set of offers can be divided into *competitive* and *noncompetitive* subsets. A sell (buy) offer at node j/m is *competitive* if nodal price π_j/π_m is higher/lower than or equal to the offer price, that is, if $c_j \leq \pi_j/e_m \geq \pi_m$. An offer j is *strictly competitive* if $c_j < \pi_j$ or $e_m > \pi_m$. A sell (buy) offer j/m is considered to be noncompetitive offer if the nodal price is lower/higher than the offer price, that is, if $c_j > \pi_k/e_m < \pi_k$. If an accepted (fully or partially) offer is also noncompetitive offer then it is called *forced* offer.

Two notions of fairness have been introduced in [26]: absolute fairness and relative fairness.

4.1. Absolute Fairness. Fairness in absolute sense means that no offer brings individual loss. Figure 2 illustrates an accepted (forced) sell offer j. Since the seller is paid with nodal price π_j lower than the offer price c_j, the entity gains loss $q_j(c_j - \pi_j)$. To

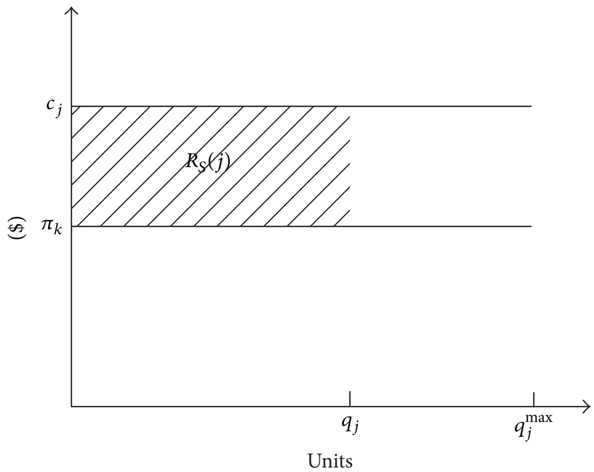


Figure 2: Cost $R_S(j)$ of forced sell.

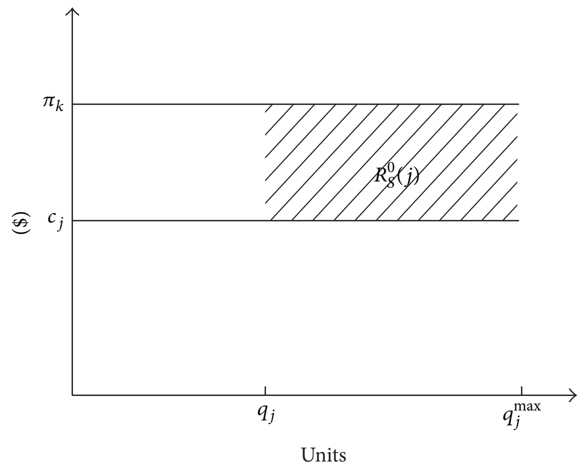


Figure 3: Loss of profits $R_S^0(j)$ for rejected, competitive sell offer.

compensate the loss, the cost of forced sell $R_S(j) = (c_j - \pi_j)q_j$ must be also paid to the seller j. Of course, if it is possible to set nodal prices, that make each offer competitive, then no individual losses appear. A mechanism $\mathscr{M}$ can ensure no individual losses either by directly setting prices or by additional payment of losses compensation $R_S(j)$.

Similarly, the compensation of competitive buyer losses should be introduced. For forced buy offer the compensation is $R_B(m) = (\pi_m - e_m)d_m$. According to the concept of fairness introduced in [26], a mechanism $\mathscr{M}$ is *fair in absolute sense* if it ensures no individual losses.

4.2. Relative Fairness. Mechanism is *fair in the relative sense* if none of the parties can claim to be treated worse than others. This means that each strictly competitive offer must be fully accepted. For instance, a seller can claim to be treated clearly unfair if his offer price is lower than market nodal price, but his offer is not winning. If such situation cannot be avoided, the compensations for the loss of profit $R_S^0(j)$ should be introduced. Figure 3 illustrates the compensation for a sell offer j which is $R_S^0(j) = (q_j^{\max} - q_j)(\pi_j - c_j)$. Similarly, the compensation for uncompetitive buy offer m is defined as $R_B^0(m) = (d_m^{\max} - d_m)(e_m - \pi_m)$.

Also, two participants connected with undersaturated edge should have the same nodal prices. If this would not be satisfied, then the entity with the worse nodal price could trade in the neighbor's node with better profits.

A mechanism $\mathscr{M}$ is fair if it satisfies fairness conditions in absolute and relative senses.

5. Analysis of Fairness

5.1. Price of Fairness. Let Q be the social welfare resulting from the optimal solution of VWDP. Let $\mathscr{M}$ be a mechanism that satisfies absolute and relative fairness conditions. The mechanism produces some welfare distribution. Some part of the social welfare is distributed under pure market rules. The sellers obtain $Q^S = \sum_{j\in\mathscr{J}} \max\{0, \pi_j - c_j\} * q_j$ and the buyers gain $Q^B = \sum_{m\in\mathscr{B}} \max\{0, e_m - \pi_m\} * d_m$. Some part of the social welfare is also distributed via lost profit compensations: $R_S^0(j)$ and $R_B^0(m)$. Finally, some social welfare may be not distributed and market becomes imbalanced.

The difference between total social welfare and welfare distributed under the pure market rules is a price that must be paid to maintain fairness [34]. It can be partially paid as the compensations or it becomes an unbalanced value that should be allocated with some additional, not market, mechanism (like fees and subsidies). We call this difference the *price of fairness* (PoF).

5.2. Space of Value Mechanisms. Let $e = (e_m)$, $c = (c_j)$, $d = (d_m)$, $q = (q_j)$, $\pi = (\pi_v)$, $R_S = (R_S(j))$, $R_B = (R_B(m))$, $R_S^0 = (R_S^0(j))$, $R_B^0 = (R_B^0(m))$. Let us define $\mathfrak{M}$ the space of value mechanisms as follows:

$$\mathfrak{M} = \left\{\mathscr{M} : (e, c, d, q) \longrightarrow \left(\pi, \pi_0, R_S, R_B, R_S^0, R_B^0\right)\right\}. \quad (4)$$

The space $\mathfrak{M}$ is quite general. It covers LMP mechanism ($R_S = R_B = R_S^0 = R_B^0 = 0$ and π are marginal). Also the single uniform price model is in $\mathfrak{M}$ (π_0 is uniform market price and $\pi_v = \pi_0, \forall v \in V$).

Moreover, for the sake of generality, we also consider the mechanisms that, unlike the LMP or classical uniform pricing, assume two nodal market prices at each node: buying price π_v^B and selling price π_v^S. The idea of differentiation of sell and buy prices has been proposed in [26] and discussed in [31] in the context of balancing electrical energy market.

5.3. PoF Minimization. In the perfect situation, when no congestion manifests in the network, the maximal economic benefit $Q^u = \sum_{m\in\mathscr{B}} e_m * d_m^u - \sum_{j\in\mathscr{J}} c_j q_j^u$ can be reached, where d_m^u and q_j^u are the accepted volumes under neglected network constraints assumption. We will call the solution $d^u = (d_m^u)$, $q^u = (q_j^u)$ of the VWDP with network constraints neglected the unconstrained solution. The related social welfare Q^u will be called the unconstrained welfare. In most cases, as a result of limited resources in the system, the loss of aggregated economic benefits is observed. We refer to the solution $d = (d_m)$, $q = (q_j)$ as constrained solution for which constrained social welfare Q is obtained. Let D be the total trade volume, $D = \sum_j q_j = \sum_j q_j^u$.

Lemma 2. *Price of fairness can be expressed as follows:*

$$PoF = \left(\pi_0^B - \pi_0^S\right) D + \sum_j \pi_{j0} q_j + \sum_i \pi_{0i} d_i. \quad (5)$$

Proof. Market sell cost received by the sellers is

$$K\left(\pi_0^S, (\pi_{j0})\right) = \sum_j \left(\pi_0^S - \pi_{j0}\right) q_j + R_S + R_S^0. \quad (6)$$

Market buy cost received by the buyers is

$$Z\left(\pi_0^B, (\pi_{0i})\right) = \sum_i \left(\pi_0^B + \pi_{0i}\right) d_i - R_B - R_B^0. \quad (7)$$

The basic balance is as follows: $K(\pi_0^S, (\pi_{j0})) + Q^0 = Z(\pi_0^B, (\pi_{0i}))$, where Q^0 is an imbalance of the market. Thus $\sum_j(\pi_0^S - \pi_{j0})q_j + R_S + R_S^0 + Q^0 = \sum_i(\pi_0^B + \pi_{0i})d_i - R_B - R_B^0$, and after transformation $Q^0 + R_S + R_B + R_S^0 + R_B^0 = \sum_i(\pi_0^B + \pi_{0i})d_i - \sum_j(\pi_0^S - \pi_{j0})q_j$, and after further simplifications $Q^0 + R_S + R_B + R_S^0 + R_B^0 = (\pi_0^B - \pi_0^S)D + \sum_i \pi_{0i} d_i + \sum_j \pi_{j0} q_j$. □

Let us analyze the properties of $K(\pi_0^S, (\pi_{j0}))$. For given values of (π_{j0}) we obtain the market sell cost function $K(\pi_0^S)$ of one variable π_0^S. If π_0^S is sufficiently small (i.e., it is below each of the offer prices), then the function $K(\pi_0^S)$ is equal to some minimal value K_0. Function $K(\pi_0^S)$ increases with π_0^S increase. $K(\pi_0^S)$ is a continuous, convex function (the derivative of this function is equal to volume of sell offers and prices lower than π_0^S).

Similar analysis can be done for function $Z(\pi_0^B, (\pi_{0i}))$. For given values of (π_{0i}) we obtain function $Z(\pi_0^B)$, which for sufficiently big price π_0^B, no less than the maximal offer price, is equal to sell value Z_0. When the value of π_0^B goes down, the function $Z(\pi_0^B)$ becomes decreasing and its derivative is increasing. So, it is continuous, concave function.

Functions $K(\pi_0^S)$ and $Z(\pi_0^B)$ are illustrated in Figure 4. It is easy to see from (5) and Figure 4 that, for given values (π_{j0}) and (π_{0i}), the PoF function of market prices π_0^S and π_0^B is a convex function. Let us denote the value of market trade by $W = K(\pi_0^S) + Q^0 = Z(\pi_0^B)$ for given values (π_{j0}) and (π_{0i}).

Lemma 3. *The PoF is a convex function with respect to the value of market trade W.*

Proof. The proof of convexity of function $(\pi^B - \pi^S)D$, without congestion costs, is provided in [26]. For fixed congestion costs, that is, for given values of (π_{j0}) and (π_{0i}), PoF(W) is a function of market trade value W in the range $[K_0, Z_0]$ and it is a convex function (see Figure 4).

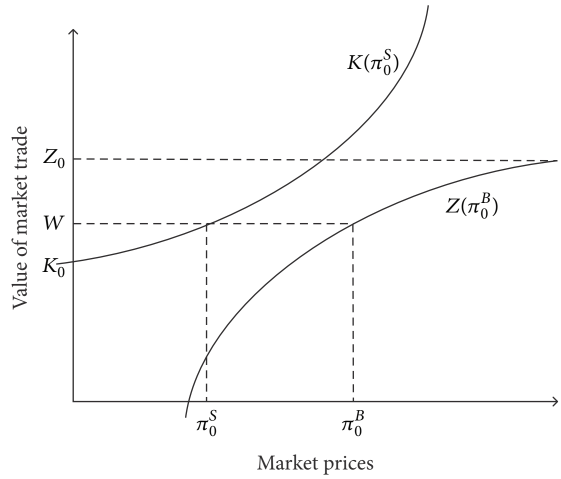


Figure 4: Market sell cost and market buy cost functions.

Notice that above property is true for all values of (π_{j0}) and (π_{0i}). Any change in values of (π_{j0}) results in parallel shift of the function $K(\pi_0^S)$ by the value $-\sum_j \pi_{j0} q_j$. Any change in values of (π_{0i}) results in parallel shift of function $Z(\pi_0^B)$ by $\sum_j \pi_{0i} d_j$. Due to market balance, any deviation in $\sum_j \pi_{j0} q_j$ must be compensated by a change in $\sum_j \pi_{0i} d_j$; thus $\sum_j \pi_{j0} q_j + \sum_j \pi_{0i} d_j$ is constant. Finally, any shift in $K(\pi_0^S)$ must also result in the same shift in $Z(\pi_0^B)$. For different prices (π_{j0}) and (π_{0i}) the gap between $K(\pi_0^S)$ and $Z(\pi_0^B)$ is constant and PoF(W) does not change when the values (π_{j0}) and (π_{0i}) are changing. Therefore, PoF(W) is a convex function for any prices (π_{j0}) and (π_{0i}). □

Now, we will show that the minimum of PoF is reached for market price $\pi_0 = \pi^u$. First, in the following two lemmas we consider two cases: (1) market price π_0 is lower than π^u; (2) market price π_0 is greater than π^u.

Lemma 4. *The PoF for market price $\pi_0 = \pi^u$ is lower than PoF for $\pi_0 < \pi^u$.*

Proof. We assume that congestions involve some reduction of accepted volume of offers from set J^R with total reduced volume $q^R = \sum_{j \in J^R}(q_j^u - q_j)$. Because these offers would be accepted in the unconstrained market, their prices must be lower than π^u. To meet the demand D, the reduction must be compensated by some forced offers. Let us denote the set of these offers by J^W. Their prices must be not lower than π^u and their total volume is $q^R = \sum_{j \in J^W}(q_j - q_j^u)$. Now, we will consider a decrease of market price by the value $\pi^u - \pi_0$. It would decrease the compensations by the value $(\pi^u - \pi_0) * q^R$, if π_0 is higher than c_j for $j \in J^R$, or the decrease would be even smaller if some of the offers would become competitive. Total change of compensation is determined as follows:

$$\sum_{j \in J^R, c_j \le \pi_0} (\pi^u - \pi_0) * \left(q_j^u - q_j\right) + \sum_{j \in J^R, c_j > \pi_0} \left(\pi^u - c_j\right) * \left(q_j^u - q_j\right). \quad (8)$$

The above expression can be also rewritten in the following form:

$$\sum_{j\in J^R} \left(\pi^u - \pi_0\right) * \left(q_j^u - q_j\right) - \sum_{j\in J^R, c_j>\pi_0} \left(c_j - \pi_0\right) * \left(q_j^u - q_j\right). \tag{9}$$

Decreasing the market price causes increase in costs of forced sell with value $(\pi^u - \pi_0) * q_j$ for $j \in J^W$ and $(c_j - \pi_0) * q_j$ for $j \in J^R$ and $c_j > \pi_0$. Total change in the PoF is as follows:

$$\begin{aligned}
&\text{increase of forced costs} \\
&\quad - \text{decrease of cost of compensations} \\
&= \left(\pi^u - \pi_0\right) * q^R + \sum_{j\in J^R, c_j>\pi_0} \left(c_j - \pi_0\right) * q_j \\
&\quad - \left(\pi^u - \pi_0\right) * q^R + \sum_{j\in J^R, c_j>\pi_0} \left(c_j - \pi_0\right) * \left(q_j^u - q_j\right) \\
&= \sum_{j\in J^R, c_j>\pi_0} \left(c_j - \pi_0\right) * q_j \\
&\quad + \sum_{j\in J^R, c_j>\pi_0} \left(c_j - \pi_0\right) * \left(q_j^u - q_j\right) > 0.
\end{aligned} \tag{10}$$

Similar reasoning can be carried out for the buyers, showing that decreasing π_0 leads to increase of the PoF. □

Lemma 5. *The PoF for market price $\pi_0 = \pi^u$ is lower than PoF for $\pi_0 > \pi^u$.*

Proof. The proof is similar to the proof of the previous lemma. Let us assume that congestions cause the reduction of accepted volume of offers from the set J^R with total volume $q^R = \sum_{j\in J^R}(q_j^u - q_j)$. Because these offers would be accepted in the unconstrained market, their prices must be not higher than π^u. To meet the demand D, there must be also forced increase of other offers. Let us denote these offers by J^W. Their prices must be not lower than π^u and their total volume is $q^R = \sum_{j\in J^W}(q_j - q_j^u)$. Increase of market price by value $\pi_0 - \pi^u$ causes a compensation increase by value

$$\begin{aligned}
&\sum_{j\in J^R} \left(\pi_0 - \pi^u\right) * \left(q_j^u - q_j\right) + \sum_{c_j=\pi^u} \left(\pi_0 - \pi^u\right) * \left(q_j^{\max}\right) \\
&\quad + \sum_{j\in J^W, c_j<\pi_0} \left(\pi^0 - c_j\right) * \left(q_j^{\max} - q_j\right) \\
&\quad + \sum_{j:q_j=0, \pi^u<c_j<\pi_0} \left(\pi^u - c_j\right) * q_j^{\max}.
\end{aligned} \tag{11}$$

The increase of market price causes the decrease cost of forced sell by

$$\sum_{j\in J^W, c_j>\pi_0} \left(\pi_0 - \pi^u\right) * q_j + \sum_{j\in J^W, c_j\le\pi_0} \left(\pi^0 - c_j\right) * q_j. \tag{12}$$

The above expression can be also rewritten as follows:

$$\left(\pi_0 - \pi^u\right) * q^R - \sum_{j\in J^W, c_j>\pi_0} \left(\pi^0 - c_j\right) * q_j. \tag{13}$$

Total change in PoF of balancing is as follows:

$$\begin{aligned}
&\text{increase of cost of compensations} \\
&\quad - \text{decrease of forced costs} \\
&= \left(\pi_0 - \pi^u\right) * q^R + \sum_{c_j=\pi^u} \left(\pi_0 - \pi^u\right) * \left(q_j^{\max}\right) \\
&\quad + \sum_{j\in J^W, c_j<\pi_0} \left(\pi^0 - c_j\right) * \left(q_j^{\max} - q_j\right) \\
&\quad + \sum_{j:q_j=0, \pi^u<c_j<\pi_0} \left(\pi^u - c_j\right) * q_j^{\max} \\
&\quad - \left(\pi_0 - \pi^u\right) * q^R + \sum_{j\in J^W, c_j>\pi_0} \left(\pi^0 - c_j\right) * q_j > 0.
\end{aligned} \tag{14}$$

Similar reasoning can be carried out for the buyers, showing that increasing π_0 leads to increase of the PoF. □

Finally, we can formulate the theorem about the minimal PoF.

Theorem 6. *The minimum of PoF is achieved at the market price π^u at every node.*

Proof. The proof is clear on the basis of Lemma 3 about the convexity of PoF and Lemmas 4 and 5 about local minimum of PoF at price π^u. □

6. Summary

In the paper, we have introduced and analyzed the concept of *price of fairness*. PoF reflects the social welfare that must be distributed out of pure market mechanism to assure that the final distribution is fair according to the absolute and relative fairness definitions. Our main result is proving that the PoF is a convex function with respect to the market value W. It means that there exists the unique minimum of PoF function. In fact, we have proved that the minimal PoF is achieved for theoretical uniform price π^u coming from the unconstrained market solution. Instead of nodal marginality, the costs of forced offers and compensations of lost profits are paid. We observe that widely applied locational marginal pricing mechanism does not in general minimize the PoF.

Our results open the doors for further investigations of new mechanisms. We have shown that there is still a lot of space for new mechanisms that can be even better than locational marginal pricing, at least in some criteria, for example, PoF. We have introduced the space of mechanism $\mathfrak{M}$, which is promising to contain new interesting mechanism designs. We believe that further researches should include more quality measures of mechanism, for example, efficiency market signals.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

References

[1] M. Kaleta, "Security constrained network winner determination problem," in *Automatyzacja Procesów Dyskretnych, Teoria i Zastosowania*, A. Świerniak and J. Krystek, Eds., pp. 111–118, Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, 2012.

[2] M. Kaleta and T. Traczyk, *Modeling Multi-Commodity Trade: Information Exchange Methods, Advances in Intelligent and Soft Computing*, Springer, 2012.

[3] H. Cremer and J. Laffont, "Competition in gas markets," *European Economic Review*, vol. 46, no. 4-5, pp. 928–935, 2002.

[4] H. Cremer, F. Gasmi, and J. Laffont, "Access to pipelines in competitive gas markets," *Journal of Regulatory Economics*, vol. 24, no. 1, pp. 5–33, 2003.

[5] A. Dinar, R. Howitt, S. Rassenti, and V. L. Smith, "Development of water markets using experimental economics," in *Markets for Water*, K. Easter, M. Rosegrant, and A. Dinar, Eds., vol. 15 of *Natural Resource Management and Policy*, pp. 259–275, Springer, New York, NY, USA, 1998.

[6] P. Kacprzak, M. Kaleta, P. Pałka, K. Smolira, E. Toczyłowski, and T. Traczyk, "Application of open multi-commodity market data model on the communication bandwidth market," *Journal of Telecommunications and Information Technology*, vol. 4, pp. 45–50, 2007.

[7] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn, *Spot Pricing of Electricity*, Kluwer Academic Publishers, New York, NY, USA, 1988.

[8] W. W. Hogan, "Contract networks for electric power transmission," *Journal of Regulatory Economics*, vol. 4, no. 3, pp. 211–242, 1992.

[9] S. Lochner, "Identification of congestion and valuation of transport infrastructures in the European natural gas market," *Energy*, vol. 36, no. 5, pp. 2483–2492, 2011.

[10] A. Rosenberg, "A proposed alternative to locational marginal pricing and its impact on new generation," *The Electricity Journal*, vol. 17, no. 3, pp. 11–25, 2004.

[11] H. Luss, "On equitable resource allocation problems: a lexicographic minimax approach," *Operations Research*, vol. 47, no. 3, pp. 361–378, 1999.

[12] W. Ogryczak, M. Pioro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problems," *Journal of Telecommunications and Information Technology*, vol. 4, pp. 43–53, 2005.

[13] H. Luss, *Equitable Resource Allocation: Models, Algorithms and Applications*, Information and Communic ation Technology Series, Wiley, 2012.

[14] H. P. Young, *Equity: In Theory and Practice*, Princeton University Press, 1995.

[15] A. B. Atkinson, "On the measurement of inequality," *Journal of Economic Theory*, vol. 2, pp. 244–263, 1970.

[16] A. Sen, *On Economic Inequality*, Clarendon Paperbacks, Clarendon Press, 1973.

[17] W. Ogryczak, "Inequality measures and equitable locations," *Annals of Operations Research*, vol. 167, pp. 61–86, 2009.

[18] O. Maimon, "The variance equity measure in locational decision theory," *Annals of Operations Research*, vol. 6, no. 5, pp. 147–160, 1986.

[19] E. Erkut, "Inequality measures for location problems," *Location Science*, vol. 1, no. 3, pp. 199–217, 1993.

[20] W. Ogryczak and T. Trzaskalik, "Equity, fairness and multicriteria optimization," in *Multiple Criteria Decision Making '05*, pp. 185–199, Wydawnictwo Uczelniane AE Katowice, 2006.

[21] M. M. Kostreva and W. Ogryczak, "Linear optimization with multiple equitable criteria," *RAIRO Operations Research*, vol. 33, no. 3, pp. 275–297, 1999.

[22] V. Krishna and V. Krishna, *Auction Theory*, Academic Press, 2002.

[23] S. Sudeendra, M. Saini, and S. Rao, "Fairness in combinatorial auctions," CoRR abs/1005.4774, 2010.

[24] J. Murillo, B. López, V. Muñoz, and D. Busquets, "Fairness in recurrent auctions with competing markets and supply fluctuations," *Computational Intelligence*, vol. 28, no. 1, pp. 24–50, 2012.

[25] C.-C. Wu, C.-C. Chang, and I.-C. Lin, "New sealed-bid electronic auction with fairness, security and efficiency," *Journal of Computer Science and Technology*, vol. 23, no. 2, pp. 253–264, 2008.

[26] E. Toczyłowski, *Optimization of Market Processes under Constraints*, EXIT Academic Publishing, 2003, (Polish).

[27] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," *Operations Research*, vol. 59, no. 1, pp. 17–31, 2011.

[28] T. Bonald and L. Massouli, "Impact of fairness on internet performance," in *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '01)*, pp. 82–91, ACM, New York, NY, USA, 2001.

[29] M. Butler and H. P. Williams, "Fairness versus efficiency in charging for the use of common facilities," *Journal of the Operational Research Society*, vol. 53, no. 12, pp. 1324–1329, 2002.

[30] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.

[31] E. Toczyłowski and I. Zoltowska, "A new pricing scheme for a multi-period pool-based electricity auction," *European Journal of Operational Research*, vol. 197, no. 3, pp. 1051–1062, 2009.

[32] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pp. 404–413, Trier, Germany, March 1999.

[33] C. Papadimitriou, "Algorithms, games, and the internet," in *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC '01)*, pp. 749–753, ACM Press, New York, NY, USA, 2001.

[34] M. Kaleta, "Computing alpha-efficient cost allocations for unbalanced games," in *Social Informatics*, vol. 6430 of *Lecture Notes in Computer Science*, pp. 103–112, 2010.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 938483, 5 pages
http://dx.doi.org/10.1155/2014/938483

Research Article

Design of Optical Wireless Networks with Fair Traffic Flows

Artur Tomaszewski

Institute of Telecommunications, Warsaw University of Technology, Poland

Correspondence should be addressed to Artur Tomaszewski; artur@tele.pw.edu.pl

Received 7 March 2014; Accepted 7 May 2014; Published 30 June 2014

Academic Editor: Dritan Nace

The paper presents a method for optimising the wireless optical network that carries elastic packet traffic. The particular focus is on modelling the effect of elastic traffic flows slowing down in response to the decrease of the optical transmission systems' capacity at bad weather conditions. A mathematical programming model of the network design problem is presented that assumes that the packet rates of elastic traffic flows decrease fairly. While practically any subset of network links can be simultaneously affected by unfavourable transmission conditions, a particular challenge of solving the problem results from a huge number of network states considered in the model. Therefore, how the problem can be solved by generating the most unfavourable network states is presented. Moreover, it is proved that it is entirely sufficient to consider only the states that correspond to the decrease of capacity on a single link. Finally, as the general problem is nonlinear, it is shown that the problem can be transformed to a linear MIP problem and solved effectively when single-path routing of traffic flows is assumed.

1. Problem Definition

The paper considers the problem of designing a wireless optical network that carries elastic packet traffic. Each node of the considered network is a packet router, while each link connecting a pair of nodes is a packet link composed of a number of optical wireless transmission systems (cf. Figure 1).

Employing optical wireless transmission systems has many advantages, the major one being that little network infrastructure is required. Thus, the network can be installed or expanded quickly and the installation process is comparatively inexpensive. The network is also very flexible as far as reconfiguration is concerned—a node can be easily reconnected to a different set of nodes and the transmission equipment can be deinstalled and moved to another location quickly. However, there are also disadvantages. For example, the transmission systems are line-of-sight systems, which means that the optical transmitter must see the optical receiver. That limits the feasible network configurations and may heavily influence the network design as shown in [1].

The most important issue, however, is that the transmitted signal is vulnerable to atmospheric conditions (smoke, fog, smog, etc.) as the systems are wireless and no transmission medium is used (cf. [2]). The worsening of atmospheric conditions may heavily influence the quality of the received optical signal. With that in mind, the optical wireless transmission systems are designed to operate in a number of transmission modes. Based on a specific scheme of signal coding and modulation, each mode provides particular capacity of the transmission system and particular robustness of the optical signal to unfavourable transmission conditions—in general, the higher the system's bit-rate the lower the signal's robustness. Thus, using multiple schemes effectively solves the trade-off between signal robustness and system capacity.

At perfect weather conditions the transmission systems of the wireless optical network are supposed to operate using the highest-capacity transmission mode, which provides the maximum capacity of links. But when the signal propagation conditions deteriorate, the operating mode of each affected transmission system changes to a more robust one and the capacity of network links decreases. Whenever the resulting capacity of a network link is less than the total nominal bandwidth of the information flows assigned to the link, packet losses are inevitable.

With elastic packet traffic, packet losses will cause packet sources to adapt (lower) their packet rate to match the available network bandwidth. This behaviour is due to the

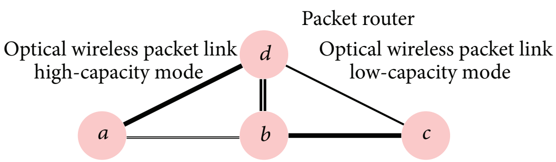


Figure 1: An optical wireless network.

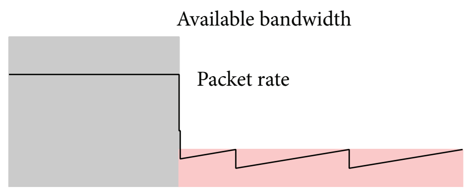


Figure 2: Elastic traffic behaviour.

per session end-to-end flow control mechanisms of the TCP data transport protocol. The packet rate of the TCP session is increased linearly every time the sender application receives an acknowledgement that a packet has reached the destination, and the rate is decreased geometrically every time a packet is lost (cf. Figure 2). Since the packet losses happen randomly, arguably, the decrease of the packet rates of the sessions on a given link must be fair, meaning that the sessions with higher packet rates decrease the rate first and all sessions decrease the rate to the same value.

Changing the rate at which packets are sent directly affects the perceived quality of service and should be controlled by careful network design. The network design problem considered in this paper consists in dimensioning the links of a given optical wireless network and routing a given set of elastic traffic flows so that the total cost of links (i.e., the total cost of the optical wireless transmission systems installed on the links) does not exceed a given budget $B > 0$, and in nonnominal transmission conditions the bandwidth that is assigned to any single traffic session is decreased by at most a given factor γ, $0 < \gamma < 1$. Alternatively, the objective of the design may be to minimise the total cost of links or to maximise the minimal bandwidth reduction. It is assumed that any subset of network links can be affected by unfavourable transmission conditions and that the decrease of packet rates on an affected link is fair.

The considered problem is similar to the classical problems of survivable network design (cf. [3]). However, there are a couple of major differences. First, in this paper a max-min fair distribution of traffic is considered instead of the general one; moreover, the fairness is considered at the level of individual sessions and not aggregated traffic flows. The application of the max-min fairness concept in network design and the approaches to solving max-min fairness problems are discussed in [4, 5]. Second, designing survivable networks is based on the notion of the set of network states, which usually consists of the nominal (failure-free) state and a number of nonnominal (failure) states. And the particular issue considered in this paper (apart from the fact that links do not fail but their capacity is decreased instead) is that in the case of the optical wireless network no meaningful set of nonnominal network states can be defined (what is the extent of fog?)—actually, any subset of links can be affected by unfavourable transmission conditions, which leads to a huge number of network states. Dealing with that issue is a major problem considered in this paper.

2. Problem Modelling

Let the optical wireless network be modelled with a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ being the set of nodes and $\mathcal{E}$ being the set of links, and let $\mathcal{P}$ be the set of paths in $\mathcal{G}$. For each $p \in \mathcal{P}$, let $E_P(e) \subseteq \mathcal{E}$ denote the set of links of path p, and, for each $e \in \mathcal{E}$, let $P_E(e) \subseteq \mathcal{P}$ denote the set of paths that use link e.

Let the links of the network be realised with one type of transmission system that can operate in two transmission modes—the primary (high-capacity) mode which is used in nominal transmission conditions and the secondary (low-capacity) mode which is used when conditions deteriorate. Without loss of generality it can be assumed that the capacity of the system is equal to 1 in the high-capacity mode and is equal to α, $0 < \alpha < 1$, in the low-capacity mode. Each link of the network can be equipped with a number of transmission systems. Let the link cost function $\xi : \mathcal{E} \mapsto \mathbb{R}_+$ define the cost of a single transmission system installed on the link.

Let $\mathcal{S}$ denote the set of network states. Let s_0 denote the nominal network state of perfect transmission conditions (when all transmission systems operate in the high-capacity mode) and let set $\mathcal{S}_f$ of nonnominal network states represent all assumed situations of unfavourable transmission conditions; thus, $\mathcal{S} \equiv \{s_0\} \cup \mathcal{S}_f$. In particular, $\mathcal{S}_f$ could be defined by assuming that at most N links of the network can be simultaneously affected by unfavourable signal transmission conditions. For each $s \in \mathcal{S}_f$, let $E_S(s) \subseteq \mathcal{E}$ be the set of affected links in nonnominal state s. It is assumed that whenever transmission conditions on a network link deteriorate, all the transmission systems installed on the link operate in the low-capacity transmission mode.

Let the network traffic be modelled with a set of demands $\mathcal{D}$. Let function $s : \mathcal{D} \mapsto \mathbb{Z}_+$ define the number of sessions that correspond to the demand, function $r : \mathcal{D} \mapsto \mathbb{R}_+$ define the nominal packet rate of a single session, and function $l : \mathcal{D} \mapsto \mathbb{R}_+$ define the average packet length for the demand (it can be noticed that $b(d) \equiv r(d)l(d)$ is the nominal bandwidth required by a single session and $h(d) \equiv r(d)l(d)s(d)$ is the volume of demand $d \in \mathcal{D}$). Finally, for each $d \in \mathcal{D}$, let $\mathcal{P}(d) \subseteq \mathcal{P}$ denote the set of admissible paths of demand d.

For each $e \in \mathcal{E}$, let variable y_e denote the number of transmission systems installed on link e. For each $d \in \mathcal{D}$ and $p \in \mathcal{P}(d)$, let variable x_{dp} denote the number of sessions of demand d assigned to path p, and, for each $d \in \mathcal{D}$, $p \in \mathcal{P}(d)$, and $s \in \mathcal{S}_f$ (it could be assumed that $E_P(p) \cap E_S(s) \neq \emptyset$), let variable z_{dps} denote the reduction of the packet rate of the sessions of demand d assigned to path p in state s (if z_{dps} equals 1, the packet rate is not reduced). The network design problem can now be formulated as a mathematical programme with the following constraints:

$$\sum_{e\in\mathscr{E}} \xi(e)\, y_e \le B \quad (1a)$$

$$\sum_{p\in\mathscr{P}(d)} x_{dp} = s(d) \quad d \in \mathscr{D} \quad (1b)$$

$$\sum_{d\in\mathscr{D}} \sum_{p\in\mathscr{P}(d)\cap P_E(e)} r(d)\, l(d)\, x_{dp} \le y_e \quad e \in \mathscr{E} \quad (1c)$$

$$\sum_{d\in\mathscr{D}} \sum_{p\in\mathscr{P}(d)\cap P_E(e)} r(d)\, l(d)\, x_{dp} z_{dps} \le \alpha y_e \quad s \in \mathscr{S}_f,\ e \in E_S(s) \quad (1d)$$

$$\gamma \le z_{dps} \le 1 \quad d \in \mathscr{D},\ p \in \mathscr{P}(d),\ s \in \mathscr{S}_f \quad (1e)$$

$$x_{dp} \in \mathbb{Z}_+ \quad d \in \mathscr{D},\ p \in \mathscr{P}(d) \quad (1f)$$

$$y_e \in \mathbb{Z}_+ \quad e \in \mathscr{E}. \quad (1g)$$

Let $y \equiv (y_e)_{e\in\mathscr{E}}$, $x \equiv (x_{dp})_{d\in\mathscr{D},p\in\mathscr{P}(d)}$, and $z \equiv (z_{dps})_{d\in\mathscr{D},p\in\mathscr{P}(d),s\in\mathscr{S}_f}$. Let $\mathscr{W}(B,\gamma,\mathscr{S}_f)$ denote the feasibility set defined by constraints (1a), (1b), (1c), (1d), (1e), (1f), and (1g) being the set of triples (y,x,z). Seemingly, to complete the formulation of the network design problem, additional constraints are still required to express the conditions that must be satisfied by variables z_{dps}; the constraints should reflect the effect of fairly slowing down the elastic traffic sessions in response to the decrease of link capacities in nonnominal states $s \in \mathscr{S}_f$. That effect can be modelled as follows.

Consider a given nonnominal state $s \in \mathscr{S}_f$. Consider vector $Z \equiv z_s$. Let $\mathscr{F}$ denote the set of affected network links, that is, $\mathscr{F} \equiv E_S(s)$, and let c be the vector of their capacities: $c \equiv (c_e \in \mathbb{R}_+ : e \in \mathscr{F}, c_e = \alpha y_e)$ (i.e., $c \equiv \alpha y|_{E_S(s)}$). For each $d \in \mathscr{D}$, $p \in \mathscr{P}(d)$, and $e \in E_P(p) \cap \mathscr{F}$, let t_{dpe} be a binary variable that equals 1 if the decrease of capacity of link e causes the decrease of the packet rate of the sessions of demand d that are routed along path p and 0 otherwise. For each $e \in \mathscr{F}$, let u_e be a binary variable that equals 1 if link e is saturated (i.e., the total flow $\sum_{d\in\mathscr{D}} \sum_{p\in\mathscr{P}(d)\cap P_E(e)} r(d)l(d)x_{dp}z_{dps}$ on the link is equal to the link's capacity) and 0 otherwise; the decrease of capacity of a link can cause the decrease of a session's packet rate only if the link becomes saturated. For each $e \in \mathscr{F}$, let variable v_e denote the common packet rate of the sessions that have their rate reduced due to the decrease of capacity of link e; v_e is also the maximum session packet rate on link e. The condition for packet rate reduction can be expressed as follows:

$$Z_{dp} = \min\left\{1, \min_{e\in E_P(p)\cap\mathscr{F}} \frac{v_e}{r(d)}\right\} \quad d \in \mathscr{D},\ p \in \mathscr{P}(d). \quad (2)$$

Altogether, the following constraints must hold (C and R are sufficiently large constants, e.g., $C \equiv \sum_{d\in\mathscr{D}} h(d)$ and $R \equiv \max_{d\in\mathscr{D}} r(d)$) which express the relation between packet rate reduction and capacity decrease of links from $\mathscr{F}$ (the constraints are based on the model of the max-min fair network flows proposed in [6]):

$$1 - Z_{dp} \le \sum_{e\in E_P(p)\cap\mathscr{F}} t_{dpe} \quad d \in \mathscr{D},\ p \in \mathscr{P}(d) \quad (3a)$$

$$t_{dpe} \le u_e \quad d \in \mathscr{D},\ p \in \mathscr{P}(d),\ e \in E_P(p) \cap \mathscr{F} \quad (3b)$$

$$0 \le c_e - \sum_{d\in\mathscr{D}} \sum_{p\in\mathscr{P}(d)\cap P_E(e)} r(d)\, l(d)\, x_{dp} Z_{dp} \le C(1 - u_e) \quad e \in \mathscr{F} \quad (3c)$$

$$0 \le v_e - r(d)\, Z_{dp} \le R(1 - t_{dpe}) \quad d \in \mathscr{D},\ p \in \mathscr{P}(d),\ e \in E_P(p) \cap \mathscr{F} \quad (3d)$$

$$R(1 - u_e) \le v_e \le R \quad e \in \mathscr{F} \quad (3e)$$

$$u_e \in \{0,1\} \quad e \in \mathscr{F} \quad (3f)$$

$$t_{dpe} \in \{0,1\} \quad d \in \mathscr{D},\ p \in \mathscr{P}(d),\ e \in E_P(p) \cap \mathscr{F} \quad (3g)$$

$$0 \le Z_{dp} \le 1 \quad d \in \mathscr{D},\ p \in \mathscr{P}(d). \quad (3h)$$

Let $Q(c,x,\mathscr{F})$ denote the feasible set defined by constraints (3a), (3b), (3c), (3d), (3e), (3f), (3g), and (3h), and let $Q_Z(c,x,\mathscr{F})$ denote the projection of $Q(c,x,\mathscr{F})$ onto the set of vectors Z. It can be shown that set $Q(c,x,\mathscr{F})$ is nonempty. Actually, for a vector Z to satisfy constraints (3a), (3b), (3c), (3d), (3e), (3f), (3g), and (3h), vector $(r(d)Z_{dp} : d \in \mathscr{D}, p \in \mathscr{P}(d), x_{dp} > 0)$ must be max-min fair. Thus finding vector $Z \in Q_Z(c,x,\mathscr{F})$ can be formulated as a max-min fair optimisation problem having vector $Z \equiv 0$ as a feasible solution.

Assuming the objective of minimising the total cost of links, the considered network design problem, denoted by $P(\gamma,\mathscr{S}_f)$, can now be defined as follows:

$$P(\gamma,\mathscr{S}_f): \quad \min B \quad (4a)$$

$$(y,x,z) \in \mathscr{W}(B,\gamma,\mathscr{S}_f) \quad (4b)$$

$$z_s \in Q_Z\left(\alpha y|_{E_S(s)}, x, E_S(s)\right) \quad s \in \mathscr{S}_f \quad (4c)$$

$$B \ge 0. \quad (4d)$$

3. Problem Solving

Potentially, the number of nonnominal network states is huge as each subset of links $\mathscr{F} \subseteq \mathscr{E}$ can correspond to a nonnominal network state. Thus, in general, the number of constraints and variables of problem (4a), (4b), (4c), and (4d) is also huge. Still, problem (4a), (4b), (4c), and (4d) can be solved with the constraint and column generation approach, by generating nonnominal network states (and thus the constraints and variables that correspond to those states) on the "as-needed" basis.

Problem (4a), (4b), (4c), and (4d) defined with a restricted set $\mathscr{S}_f$ of nonnominal states can be treated as the master problem. Let (y^*, x^*, z^*) be an optimal solution to the master problem. For given y^* and x^*, the slave problem consists in finding such set of links affected by unfavourable signal transmission conditions that for at least one path of a demand the reduction of the packet rate is less than γ; this set corresponds to a new nonnominal network state s^*, which is then added to set $\mathcal{S}_f$ of the master problem.

For each $e \in \mathcal{E}$, let q_e be a binary variable that equals 1 if link e belongs to the required set $E_S(s^*)$ and 0 otherwise. For each $d \in \mathcal{D}$ and $p \in \mathcal{P}(d)$, let variable Z_{dp} denote the reduction of the packet rate of the sessions of demand d assigned to path p in the required state, and let $Z \equiv (Z_{dp})_{d\in\mathcal{D},p\in\mathcal{P}(d)}$. Let $\theta \equiv \min_{d\in\mathcal{D},p\in\mathcal{P}(d)} Z_{dp}$, and for each $d \in \mathcal{D}$ and $p \in \mathcal{P}(d)$, let r_{dp} be a binary variable that equals 1 whenever $Z_{dp} = \theta$. Then, a new nonnominal network state can be found by solving the following slave problem $R(y, x)$ for y^* and x^* which minimises the value of θ over all sets $\mathcal{F} \subseteq \mathcal{E}$ (it could also be assumed that $|\mathcal{F}| \le N$):

$$R(y,x): \quad \min \theta \tag{5a}$$

$$0 \le Z_{dp} - \theta \le 1 - r_{dp} \quad d \in \mathcal{D},\ p \in \mathcal{P}(d) \tag{5b}$$

$$\sum_{d\in\mathcal{D},p\in\mathcal{P}(d)} r_{dp} \ge 1 \tag{5c}$$

$$c_e = y_e - q_e(1-\alpha)\, y_e \quad e \in \mathcal{E} \tag{5d}$$

$$Z \in Q_Z(c, x, \mathcal{E}) \tag{5e}$$

$$c_e \ge 0 \quad e \in \mathcal{E} \tag{5f}$$

$$q_e \in \{0,1\} \quad e \in \mathcal{E} \tag{5g}$$

$$r_{dp} \in \{0,1\} \quad d \in \mathcal{D},\ p \in \mathcal{P}(d) \tag{5h}$$

$$0 \le Z_{dp} \le 1 \quad d \in \mathcal{D},\ p \in \mathcal{P}(d) \tag{5i}$$

$$0 \le \theta \le 1. \tag{5j}$$

It can be noticed that constraints (3a), (3b), (3c), (3d), (3e), (3f), (3g), and (3h) defining set $Q_Z(c, x, \mathcal{E})$ can be directly embedded into the formulation of $R(y, x)$.

If the optimal objective value θ^* of slave problem $R(y^*, x^*)$ is greater or equal to γ, the optimal solution of the master problem is the optimal solution of the considered network design problem. Otherwise, the master problem must be modified by adding variables and constraints that correspond to a new nonnominal network state $s^* \in \mathcal{S}_f$: the optimal solution vector $q^* \equiv (q_e^*)_{e\in\mathcal{E}}$ of problem (5a), (5b), (5c), (5d), (5e), (5f), (5g), (5h), (5i), and (5j) defines the values of the characteristic function of set $E_S(s^*)$; that is, $E_S(s^*) \equiv \{e \in \mathcal{E} : q_e^* = 1\}$.

Solving problem (4a), (4b), (4c), and (4d) with state generation approach does not guarantee that considering a large number of network states can be avoided. However, careful examination of the packet rate reduction condition (2) leads to a conclusion that, actually, it is sufficient to consider only single-link states in the design problem. This fact can be expressed formally with the following proposition.

Proposition 1. *Consider an arbitrary set of nonnominal network states $\mathcal{S}_f$ and set $\widehat{\mathcal{S}}_f$ such that $|E_S(s)| = 1$ for each $s \in \widehat{\mathcal{S}}_f$ and $\bigcup_{s\in\mathcal{S}_f} E_S(s) = \bigcup_{s\in\widehat{\mathcal{S}}_f} E_S(s)$. If (y^*, x^*, z^*) is a feasible solution of problem $P(\gamma, \widehat{\mathcal{S}}_f)$, then there exists a feasible solution (y^*, x^*, z) of problem $P(\gamma, \mathcal{S}_f)$.*

Proof. Consider a set of vectors z_s, $s \in \mathcal{S}_f$, such that $z_s \in Q_Z(\alpha y^*|_{E_S(s)}, x^*, E_S(s))$ (recall that sets Q_Z are not empty). It will be proved that (y^*, x^*, z) is a feasible solution of problem $P(\gamma, \mathcal{S}_f)$.

Due to (4a), (4b), (4c), and (4d), it must be shown that $(y^*, x^*, z) \in \mathcal{W}(B, \gamma, \mathcal{S}_f)$. As (y^*, x^*, z^*) is a feasible solution of problem $P(\gamma, \widehat{\mathcal{S}}_f)$, therefore $(y^*, x^*, z^*) \in \mathcal{W}(B, \gamma, \mathcal{S}_f)$, and it is sufficient to prove that vector z satisfies constraints (1e); that is, $z_{dps} \ge \gamma$ for each $d \in \mathcal{D}$, $p \in \mathcal{P}(d)$, $s \in \mathcal{S}_f$.

Assume that there exist $d' \in \mathcal{D}$, $p' \in \mathcal{P}(d')$, and $s' \in \mathcal{S}_f$, such that $z_{d'p's'} < \gamma$. Due to (2), there must exist link $e' \in E_P(p) \cap E_S(s')$ such that $z_{d'p's'} = v'_{e'}/r(d')$ for $Q(\alpha y^*|_{E_S(s')}, x^*, E_S(s'))$; obviously, that link must be saturated in s'; that is, $\sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}(d)\cap P_E(e')} r(d) l(d) x_{dp} z_{dps'} = \alpha y^*_{e'}$. According to the assumption, there exists state $\widehat{s} \in \widehat{\mathcal{S}}_f$, such that $E_S(\widehat{s}) = \{e'\}$. With (y^*, x^*, z^*) being a feasible solution of problem $P(\gamma, \widehat{\mathcal{S}}_f)$, due to (4a), (4b), (4c), and (4d), $z^*_{\widehat{s}} \in Q_Z(\alpha y^*|_{E_S(\widehat{s})}, x^*, E_S(\widehat{s}))$ and $z^*_{dp\widehat{s}} \ge \gamma$ for each $d \in \mathcal{D}$ and $p \in \mathcal{P}(d)$, in particular, $z^*_{d'p'\widehat{s}} \ge \gamma$. Therefore, due to (2), $\widehat{v}_{e'}/r(d') \ge \gamma$ for $Q(\alpha y^*|_{E_S(\widehat{s})}, x^*, E_S(\widehat{s}))$, and thus $\widehat{v}_{e'} > v'_{e'}$. Then, once again due to (2) (and the fact that $\mathcal{F} = E_S(\widehat{s}) = \{e'\}$), $z^*_{dp\widehat{s}} = \min\{1, \widehat{v}_{e'}/r(d)\} \ge \min\{1, v'_{e'}/r(d)\} \ge z_{dps'}$ for each $d \in \mathcal{D}$, $p \in \mathcal{P}(d) \cap P_E(e')$, and $z^*_{d'p'\widehat{s}} > z_{d'p's'}$. Thus, $\sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}(d)\cap P_E(e')} r(d) l(d) x_{dp} z^*_{dp\widehat{s}} > \sum_{d\in\mathcal{D}} \sum_{p\in\mathcal{P}(d)\cap P_E(e')} r(d) l(d) x_{dp} z_{dps'} = \alpha y^*_{e'}$, which is not possible. □

The intuition behind the proof of Proposition 1 is that the packet rate on any given link decreases most if that link is the only overloaded link in the network; decreasing the capacity of other links can only decrease the load on that particular link and minimise the need for packet rate decrease. Consequently, the opposite of Proposition 1 is not true as stated by the following proposition.

Proposition 2. *Consider an arbitrary set of nonnominal network states $\mathcal{S}_f$ and set $\widehat{\mathcal{S}}_f$ such that $|E_S(s)| = 1$ for each $s \in \widehat{\mathcal{S}}_f$ and $\bigcup_{s\in\mathcal{S}_f} E_S(s) = \bigcup_{s\in\widehat{\mathcal{S}}_f} E_S(s)$. If (y^*, x^*, z^*) is a feasible solution of problem $P(\gamma, \mathcal{S}_f)$, then there need not exist a feasible solution (y^*, x^*, z) of problem $P(\gamma, \widehat{\mathcal{S}}_f)$.*

Proof. Consider a network illustrated in Figure 3, which consists of nodes a, b and c, links $e_1 \equiv (a, b)$ and $e_2 \equiv (b, c)$, and demands d_1 from a to b and d_2 from a to c, such that $h(d_1) = 1/2$, $r(d_1) = 1$ and $h(d_2) = 1$, $r(d_2) = 1/2$. Let $\mathcal{S}_f \equiv \{s\}$, such that $E_S(s) \equiv \{e_1, e_2\}$. Let $\alpha = 1/2$ and $\gamma = 1/2$. Consider a solution to problem $P(\gamma, \mathcal{S}_f)$ (obviously, it is the optimal solution), such that $y^*_{e_1} = 2$ and $y^*_{e_2} = 1$

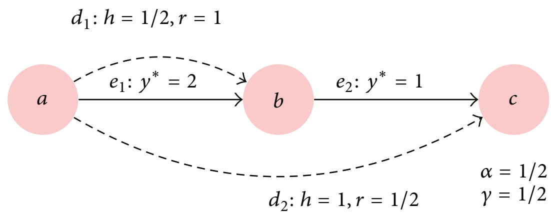


FIGURE 3: A network example for infeasibility of problem $P(\gamma, \widehat{\mathcal{S}}_f)$.

and $x^*_{d_1} = s(d_1)$ and $x^*_{d_2} = s(d_2)$ (index p is skipped as there is only one path for each demand): it can be checked that $z^*_{d_1 s} = 1$ and $z^*_{d_2 s} = 1/2$ (again with index p skipped). However, for state $\widehat{s}$, such that $E_S(\widehat{s}) \equiv \{e_1\}$, link e_1 is saturated since $h(d_1) + h(d_2) > \alpha y^*_{e_1} = 1$, and inevitably the packet rates of both flows must decrease as $r(d_1) > r(d_2)$ and $h(d_2) = \alpha y^*_{e_1}$. Thus, $s(d_1)l(d_1)v_{\widehat{s}e_1} + s(d_2)l(d_2)v_{\widehat{s}e_1} = \alpha y^*_{e_1} = 1$ and $v_{\widehat{s}e_1} = 1/(h(d_1)/r(d_1) + h(d_2)/r(d_2)) = 2/5$, and then $z_{d_1\widehat{s}} = v_{\widehat{s}e_1}/r(d_1) = 2/5 < \gamma$. □

Proposition 1 leads directly to the following conclusion.

Corollary 3. *Consider an arbitrary set of nonnominal network states $\mathcal{S}_f$ and set $\widehat{\mathcal{S}}_f \subseteq \mathcal{S}_f$ such that $|E_S(s)| = 1$ for each $s \in \widehat{\mathcal{S}}_f$ and $\bigcup_{s\in\mathcal{S}_f} E_S(s) = \bigcup_{s\in\widehat{\mathcal{S}}_f} E_S(s)$. The optimal objective function value of problem $P(\gamma, \mathcal{S}_f)$ is equal to the optimal objective function value of problem $P(\gamma, \widehat{\mathcal{S}}_f)$.*

Proof. Due to Proposition 1, the optimal objective function value of problem $P(\gamma, \widehat{\mathcal{S}}_f)$ is an upper bound on the optimal objective function value of problem $P(\gamma, \mathcal{S}_f)$. But it is also a lower bound since problem $P(\gamma, \widehat{\mathcal{S}}_f)$ is a relaxation of problem $P(\gamma, \mathcal{S}_f)$ (due to the fact that $\widehat{\mathcal{S}}_f \subseteq \mathcal{S}_f$). □

Proposition 1 and Corollary 3 not only enable considering a limited set of network states, but also allow for simplifying constraints (3a), (3b), (3c), (3d), (3e), (3f), (3g), and (3h) defining set $Q_Z(c, x, \mathcal{F})$, due to the fact that only one link for each state needs to be analysed. First, since $|\mathcal{F}| = 1$, the number of variables and constraints is heavily reduced, and index $e \in \mathcal{F}$ is actually not required. Second, as it can be noticed that it is not necessary to examine directly whether the considered link is saturated (the link must be saturated if at least one Z_{dp} is less than 1), variables u_e and constraints (3b) and (3c) are actually not required.

Still, problem (4a), (4b), (4c), and (4d) is nonlinear due to the nonlinearity of constraint (1d). However, if single-path routing of demands is assumed, the problem can be formulated as a MIP. For each $d \in \mathcal{D}$ and $p \in \mathcal{P}(d)$, let x'_{dp} be a binary variable that equals 1 if the sessions of demand d use path p and 0 otherwise. Then variables x_{dp} can be replaced with variables x'_{dp} using substitution $x_{dp} \equiv s(d)x'_{dp}$. Finally, expression $X_{dps} \equiv x'_{dp} z_{dps}$ can be linearised requiring that $0 \le X_{dps} \le x'_{dp}$ and $z_{dps} - (1 - x'_{dp}) \le X_{dps} \le z_{dps}$.

Thus, problem $P(\gamma, \mathcal{S}_f)$ can be treated as a single-path routing MIP problem with additional constraints. And as analysed in [5, 7] the single-path routing problem can be solved quite efficiently. Hopefully, the additional constraints are not very demanding, as, in particular, the actual values of variables z_{dps} are not critical—it is only important if they are greater or equal to γ. However, the numerical experiments that are supposed to illustrate the computational complexity of the problem will be the subject of a separate paper.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by National Science Centre, Poland, under Grant 2011/01/B/ST7/02967.

References

[1] Y. Li, M. Pioro, and V. Angelakis, "Design of cellular backhaul topology using the FSO technology," in *Proceedings of the 2nd International Workshop on Optical Wireless Communications (IWOW '13)*, pp. 6–10, Newcastle Upon Tyne, UK, October 2013.

[2] D. Borah, A. Boucouvalas, C. Davis, S. Hranilovic, and K. Yiannopoulos, "A review of communication-oriented optical wireless systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, article 91, 2012.

[3] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*, Morgan Kaufman, Boston, Mass, USA, 2004.

[4] W. Ogryczak, M. Pióro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problem," *Journal of Telecommunications and Information Technology*, vol. 3, pp. 43–56, 2005.

[5] A. Tomaszewski, *Optimisation of Telecommunications Transport Networks*, Wydawnictwa Politechniki Warszawskiej, Warsaw, Poland, 2013.

[6] A. Tomaszewski, "A polynomial algorithm for solving a general max-min fairness problem," *European Transactions on Telecommunications*, vol. 16, no. 3, pp. 233–240, 2005.

[7] M. Pióro, A. Tomaszewski, M. Dzida, M. Mycek, and M. Zagożdżon, "Literature survey, models, polyhedral results, and exact and heuristic methods for shortestpath routing problems," Tech. Rep., Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland, 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 693212, 11 pages
http://dx.doi.org/10.1155/2014/693212

Research Article

Max-Min Fair Link Quality in WSN Based on SINR

Ada Gogu,[1] Dritan Nace,[2] Supriyo Chatterjea,[3] and Arta Dilo[3]

[1] *Polytechnic University of Tirana, 1010 Tirana, Albania*
[2] *Université de Technologie de Compiègne, 60205 Compiègne Cedex, France*
[3] *University of Twente, 7500 AE Enschede, The Netherlands*

Correspondence should be addressed to Ada Gogu; agogu@fti.edu.al

Received 5 March 2014; Revised 21 May 2014; Accepted 23 May 2014; Published 23 June 2014

Academic Editor: Michał Pióro

This paper addresses first the problem of max-min fair (MMF) link transmissions in wireless sensor networks (WSNs) and in a second stage studies the joint link scheduling and transmission power assignment problem. Given a set of concurrently transmitting links, the MMF link transmission problem looks for transmission powers of nodes such that the signal-to-interference and noise ratio (SINR) values of active links satisfy max-min fairness property. By guaranteeing a "fair" transmission medium (in terms of SINR), other network requirements may be directly affected, such as the schedule length, the throughput (number of concurrent links in a time slot), and energy savings. Hence, the whole problem seeks to find a feasible schedule and a power assignment scheme such that the schedule length is minimized and the concurrent transmissions have a fair quality in terms of SINR. The focus of this study falls on the transmission power control strategy, which ensures that every node that is transmitting in the network chooses a transmission power that will minimally affect the other concurrent transmissions and, even more, achieves MMF SINR values of concurrent link transmissions. We show that this strategy may have an impact on reducing the network time schedule.

1. Introduction

Wireless sensor networks (WSNs) are presently used in a wide range of applications. They are usually deployed as a standalone system or as part of a larger, more sophisticated system, such as the Internet-of-Things. However, this technology has stringent requirements mainly related to energy and wireless transmission medium. The power allocated to sensor nodes in a network is fundamentally constrained; nonetheless they have to transmit their data which costs sufficient energy. Because of interference, the concurrent wireless transmissions may be easily corrupted, which increases the number of packet retransmissions. This may cause energy depletion and delays in the network. Existing solutions such as transmission power control and blacklisting PCBL [1], adaptive transmission power control ATPC [2], and adaptive and robust topology control ART [3] use parameters like received signal strength indicator (RSSI) or packet receive rate (PRR) to evaluate the quality of a link. However, RSSI or PRR may not capture the effect of interference in particular scenarios as will be further detailed.

In addition, we consider the SINR parameter and solve the problem of fair SINR link transmission. Hence, for a set of activated links, the problem seeks to find the power of the transmitting nodes such that the SINR value is fair at each receiver. Fairness is a key consideration in WSN scenarios in order to maintain a balanced view of the network and, in this case, to give the same priority to each of the concurrent transmissions. This increases the number of potential concurrent links scheduled in the same time slot and therefore reduces the schedule length. However, the feasibility of this problem is tightly coupled with the given set of activated links and therefore the scheduling. This is a typical example that shows how the optimization problems in WSN may often lead to cross-layer ones. The cross-layer optimization problem comes to be the joint link scheduling and power assignment (JLSPA) that seeks to find an efficient link scheduling scheme, in which the power of sender nodes is set variable, such that certain requirements are satisfied.

In a time-driven WSN, sensor nodes need to send their data periodically according to a regular traffic pattern. This period is usually known as a round of data gathering.

The time needed to perform a round will be determined by the schedule length which, on the other hand, is constrained by the interference effect. The relation between these two parameters is further detailed in Sections 2 and 3.1. In this work, interference is taken into account by considering the SINR parameter. Hence the subject of our research is (i) to find a power allocation scheme which guarantees a max-min fair SINR and (ii) to solve the JLSPA problem. We design a solution for two scenarios that are slightly different: the transmission power of a node changes per slot within a frame; the power is fixed (the same) in a frame level.

The paper is organized as follows. In Section 2 we recall first the interference definition and the related SINR constraint, and we present a review of related works regarding the JLSPA problem and the power assignment schemes. The fair SINR link transmission problem is introduced in Section 3. Here we state in more detail the research motivation, the problem definition, and the mathematical model. In Section 4 we present an (centralized) approach for solving the MMF SINR link transmission problem. Next, we discuss a variant of this problem with constant powers in Section 5. In Section 6 the JLSPA problem is considered and numerical results are presented to show the performance of the method discussed in the previous section. Finally, we conclude this work in Section 7.

2. Related Work

Application of optimization theory to the design of WSN algorithms is addressed in different works and for a summary we refer the reader to [4]. Regarding MMF formulation of the problems, they are generally employed to ensure fairness related to link rates [5, 6] or end-to-end flows [7, 8]. MMF is applicable in numerous areas where it is desirable to achieve an equitable distribution of certain resources shared by competing demands and is therefore closely related to max-min or min-max optimization problems [9]. In this paper, our focus is on the SINR max-min fairness. Generally speaking, a vector of transmission links is said to be MMF with respect to SINR if the corresponding vector of SINR values is MMF; that is, one cannot increase the SINR value of some transmission link without decreasing the SINR value of some other links with lower SINR.

As far as the scheduling problem apart is strongly concerned with interference avoidance for achieving successful multiple concurrent transmissions, the key point is the interference definition. Two basic definitions can be distinguished: the protocol and the physical one. The protocol definition of interference assumes that two links, which are less than k hops ($k \geq 1$) away from each other, interfere potentially and cannot be scheduled in the same time slot. The indicated number of hops refers to the number of hops between the sender nodes of these links. On the other hand, the physical definition is based on the SINR constraint where the transmission links that do not satisfy the SINR constraint cannot be scheduled simultaneously (in the following we will use interchangeably the terms transmission links and links).

However, as we will later see, the problem remains $\mathcal{NP}$-hard regardless of the interference definition. Following the protocol definition of the SINR, the JLSPA seeks to find a minimum-length schedule for all nodes in the network such that they do not interfere with each other. In the simple case ($k = 1$), the constraint requires that two edges in the same time slot do not share a node. The scheduling problem is widely modeled as the well-known optimization problem of graph coloring in which one seeks to find the minimal number of colors (chromatic number) necessary to color a graph such that no two adjacent nodes (two nodes are considered adjacent if they are "k" hops away from each other) have the same color. Finding the chromatic number in a graph is $\mathcal{NP}$-hard; therefore different methods have been proposed for this problem by the Operation Research Community. These methods have been adapted and implemented for WSN; see, for instance, [10–13].

As the protocol model underestimates the number of successful multiple concurrent transmissions, different works [1, 14, 15] consider the cumulative interference proposed by the SINR model. For solving the link scheduling problem under physical interference, one approach seeks to classify the links according to their length (or distance). Hence, in [15–17] each link belongs to a class C_k if its length l_i is $2^k \leq |l_i| < 2^{k+1}$. The idea behind the partition of links in classes is to schedule at the same time the links with the same length or very different one. The complexity of the proposed algorithm is $O(\log^4 n)$ where n represents the number of nodes in the network. Kesselheim [18] identifies another condition that the set of links scheduled in the same time slot should satisfy. Given two links, this condition is related to the ratio between the distance of one link and the respective distance between the transmitter of this link and the receiver of the other one. The scheduling algorithm is a greedy one, which ensures that the condition is satisfied if another link is added to the set of links scheduled in a given time slot. Goussevskaia et al. [16] design greedy heuristic scheduling. For each class C_k, they partition the network area into squares of side length $a \cdot 2^k$, where a is a constant, and color the squares using 4 colors. Next, for each color they pick up the links having their receiver in different squares and assign them to a time slot. The process is repeated till all the links are scheduled. The same idea is revisited in [17]. In addition, they consider the case when the links have different demands to satisfy. Moreover, the length of the time slots is not fixed and the links may be scheduled more than once in a frame. In order to identify the links that can be scheduled simultaneously, [17] proposed a link classification based on the signal to noise ratio (SNR) value. Instead of using the SINR threshold, Santi et al. [17] consider a graded SINR model which relates the PRR with SINR values as in Figure 1. Their algorithm computes a schedule length of $O(r \cdot t)$, where r is the maximal number of receivers in a cell and t is the time needed to transmit with the minimal SINR estimated in the frame (the time for transmitting a data unit is $t = 1/f$ where f is the data rate computed according to Shanon's channel capacity formula $f = B \cdot \log_2(1 + \text{SINR})$, where B is the channel bandwidth). To identify the set of concurrent

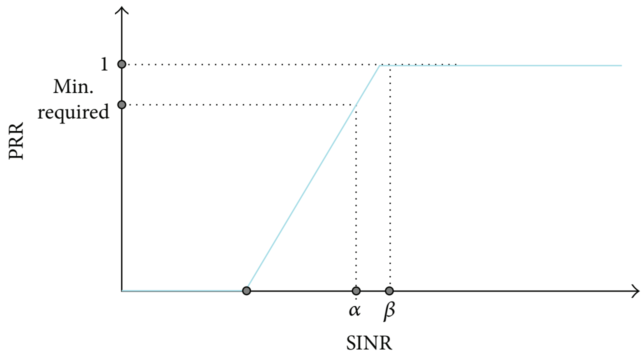


Figure 1: SINR graded model. α and β are the lower and upper bounds of the desired SINR value, respectively, as described in (2) and (4).

links, another class of algorithms solves the maximum link matching problem which seeks to find the maximum number of links in a given graph that do not have a node in common [19].

Despite the existence of different approaches for solving the scheduling problem under SINR constraint, designing a network protocol that takes this constraint into consideration is not trivial.

Till now, we have discussed the scheduling problem without emphasizing the power assignment strategy, which is of paramount importance. Based on the above discussed problems we classify these strategies into three main groups.

Uniform Power Assignment. This is the simplest and the most intuitive case where it is assumed that all the nodes use the same power to transmit their data. Hence the question is to find the optimal transmission range that maximizes the number of multiple successful concurrent transmissions [20].

Linear Power Assignment. This scheme uses the rule of assigning the power proportionally [17] to the signal attenuation (the simplest model of signal attenuation is given by d^{γ} where d is the length of the link and γ the path loss exponent) which corresponds to the minimum power of transmission that guarantees a successful packet decoding from the receiver part. Similarly, the square-root power assignments, proposed by [21], assign the power proportionally to $\sqrt{d^{\gamma}}$. The linear and the uniform strategies are frequently used in MAC layer protocols.

Nonlinear Power Assignment. According to this strategy the power is disproportional to the link distance. The study of Moscibroda and Wattenhofer [15] shows that the uniform and linear power assignment may lead to inefficient scheduling as the shortest links may "suffer" due to the high power signals emitted by the sources of the longest ones. Hence, in [15] a nonlinear power assignment strategy which gives priority to the short links is proposed. It assigns a minimal power to the longest links such that the communication is feasible and next it increases with a scaling factor the power assigned to the shorter links. Using a different scaling factor, the same scheme is applied also in [18].

This work follows on from some recent work on power assignment [22]. In this study we go further and investigate max-min fair link quality among the active transmission links.

Related to the complexity of the problem of joint scheduling and power assignments, different variants of the problem are considered. Let us first refer to the link assignment problem. This problem needs to assign the links to different time slots such that two adjacent links will not be in the same time slot and the SINR constraints will be met for each of them. It is shown in [23] that this problem is at least as hard as the edge coloring problem and is thus $\mathcal{NP}$-hard. Further, the problem of determining a minimum-length schedule that satisfies the SINR constraints is studied in [16]. By constructing a geometric instance of the scheduling problem, Goussevskaia et al. [16] show that the problem is reducible to the partition problem (given a set of integers, the partition problem seeks to decide whether it is possible to divide this set into two subsets, such that the sums of the numbers in each subset are equal). The case when the schedule has to satisfy the links demands (or flow rates) is shown to be $\mathcal{NP}$-hard by reducing it to the matching problem [24]. Hence, different variants of this problem and their respective complexities are discussed in the literature. The proof of the complexity of our JLSPA problem is presented in the work of Katz et al. [25]. The JLSPA problem was shown to be $\mathcal{NP}$-hard by [25]; when the network is embedded in the Euclidean plane, the power is variable and there are known upper and lower bounds on the power levels that can be used. Moreover, the proof remains true even for the case in which the sender node may choose its transmission power from an available set of discrete values.

3. Problem Definition

3.1. Research Motivation. For getting insights into modeling the power assignment problem under SINR constraints, we refer to some experimental test provided in [1–3]. Considering only one transmitter node and one receiver node and factoring out the issue of interference, it can be stated that the link quality between the transmitter and receiver improves as the transmitter increases its transmission power. This reasoning also holds for very sparse network where the transmitter is only within the range of the intended recipient but not in the range of any other nodes that may be simultaneously receiving data from other transmitting nodes. Under these circumstances, existing transmission power control protocols, such as those found in PCBL [1], ATPC [2], and ART [3], could help to ensure that an appropriate transmission power is used to achieve reliable link quality using minimum energy. These schemes generally use parameters such as the received signal strength indicator (RSSI) or packet receive rate (PRR) to evaluate the quality of a link. This information is then used to take decisions about the transmission power that should be set to maximize link quality using the least amount of energy. However, as the network density increases and every transmitting node is potentially within range of multiple receivers, interference plays a much larger role. Under such circumstances parameters such as RSSI or PRR do not give a good indication of whether link quality is poor or more importantly why it is poor (both power transmission and interference can be the cause). For example, if we assume that interference does not exist, higher RSSI reading generally translates into a higher PRR [1]. However, as interference increases, a higher RSSI may not result in a higher PRR, as the increased RSSI may be due to other nodes that are transmitting simultaneously and are within the range of the receiver. In addition, techniques like PCBL, ATPC, and ART only depend on information available locally at a node to make deductions about the quality of a link. As it can be seen from the performance of ART, a node may not always be able to accurately differentiate between packet loss due to a weak signal and that due to interference by using a localized approach. But as all nodes act independently of each other, one of drawbacks of such schemes is that nodes try to outdo each other. This results in higher power consumption and also has a detrimental effect on link quality. Due to these reasons, in this study we aim to find an optimal solution for transmission power assignment in a fair manner, using a centralized approach. Each node that is actively transmitting in the network chooses a transmission power that minimizes the interference effects on all the nondestination receivers. Our scheme aims to optimize the SINR parameter instead of only addressing RSSI or PRR as it is able to capture information about both the signal strength and interference more accurately.

3.2. Notation and Problem Definition. We model the wireless sensor network through a directed graph $G(V, E)$ where V is the set of nodes representing the sensors and E is the set of links representing the wireless channel communication between the sensors. For each link $(i, j) \in E$, i indicates the transmitter node and j the receiver one. The weight of link (i, j) is denoted by ω_{ij} and represents the attenuation of the signal. In some other context, ω_{ij} may be referred to as gain if it would present the signal amplification to reach the receiver. We now assume that, in a given time, only a subset of links M $(M \subset E)$ is activated. Let us denote by TX_M the subset of V containing the heads (transmitting nodes) of the directed links $(i, j) \in M$ and by RX_M the tails (receivers nodes) of links in M.

Two properties can be noticed for $\mathrm{RX}_M/\mathrm{TX}_M$ subsets:

(1) $\mathrm{RX}_M \bigcap \mathrm{TX}_M = \varnothing$,

(2) $\mathrm{RX}_M \bigcup \mathrm{TX}_M \subseteq V$.

The SINR_j value estimated in the receiver j according to [26], where $(i, j) \in M$, is given by

$$\mathrm{SINR}_j = \frac{P_i/\omega_{ij}}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a}, \quad (1)$$

where $j \in \mathrm{RX}$, P_i and P_k are the power assigned to the sender nodes, ω_{ij} denotes the weight of the transmission link (i, j), P_k/ω_{kj} measures the interference of the other links over the receiver node j of the link (i, j), where $(i, j) \in M$, and N_a is the floor noise which is considered as constant. Next, we define explicitly the parameters of a successful transmission. Clearly, a crucial parameter for estimating the link quality is PRR. This parameter is strongly related to SINR [17, 27]. According to these works, the packets are successfully received only when SINR exceeds a given threshold. The graded SINR model graphically presented in Figure 1 shows the relation between PRR and SINR which is used in this work.

3.3. Mathematical Modeling. Given a set of concurrently transmitting links, the max-min fair link transmissions problem determines the transmission power allocated to nodes such that the SINR values of active links are MMF. For solving this problem, we will refer to a subproblem which is modeled as max-min linear programming. Let us present in detail the constraints and the objective function for this subproblem. First, in order to have fair link transmissions, we aim to maximize the minimum SINR value associated with receiver nodes. Moreover, this objective permits improving the quality of the worst link which usually comes out to be a key point for measuring the network performance. Let us have a look at the constraints.

(1) To have a successful transmission we require the SINR value at the receiver to be bigger than a threshold. We denote by α this lower bound as given in

$$\mathrm{SINR}_j = \frac{P_i/\omega_{ij}}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a} \geq \alpha, \quad j \in \mathrm{RX}_M. \quad (2)$$

(2) The intended signal strength measured by received signal strength indicator (RSSI) has to be bigger than a threshold RSSI_0. This threshold represents the lowest power level of the signal which permits a receiver to detect and decode the information of the signal. It is also known as the receiver sensitivity and can be easily found in the data sheet of the radio transceiver:

$$\mathrm{RSSI}_{ij} = \frac{P_i}{\omega_{ij}} \geq \mathrm{RSSI}_0, \quad (i, j) \in M, \quad (3)$$

where RSSI_{ij} is the received strength indicator at the node j when the link $(i, j) \in M$ is activated.

(3) Considering the graded SINR model in Figure 1, we can observe that beyond a given threshold β of SINR, the PRR does not change. It is reasonable to keep the SINR values as close as possible to the β value. Imposing β as an upper bound for the SINR of all receivers has the high risk of infeasible solutions. Instead, we add the constraint for the lowest SINR as follows:

$$\min_{j\in \mathrm{RX}} \mathrm{SINR}_j \leq \beta. \quad (4)$$

4. Max-Min Fair SINR Link Transmission

We investigate in this section the problem of max-min fair SINR link transmission (MMFSLT). With respect to

Input: Set of activated links $M = (\text{TX}, \text{RX})$;
Output: $S = [s_{(l,p)} : (l, p) \in M]$ the vector of optimal Max-Min Fair SINR associated with activated links
(1) Set $L = M$; $L_0 = \emptyset$; $k = 1$;
(2) **repeat**
(3) Solve problem D_k (Compute z value);
(4) Identify $(l, p) \in L$ for the respective z value;
(5) Set $s_{(l,p)} = z$;
(6) Set $L_k = L_{k-1} \cup (l, p)$;
(7) Set $L = L \setminus (l, p)$;
(8) $k = k + 1$;
(9) **until** $((z \leq \beta)$ **AND** $(L$ *not empty*$))$;
(10) For all remaining links (l, p) after the final step of the algorithm, set their respective $s_{(l,p)} = \beta$;

Algorithm 1: MMF algorithm (optimal max-min fair SINR link transmission).

Input: M the set of activated links, TX nodes (indexed by i), RX nodes (indexed by j);
Output: z value;
(1) $z := \alpha$;
(2) **while** $((\epsilon > 0)$ *and* $(z \leq \beta))$ **do**
(3) Solve D_1' (Compute ϵ and P_i values);
(4) **for all the** $(i, j) \in M$ **do**
(5) $z \leftarrow z + \min_{(i,j)\in M} \left\{ \dfrac{\epsilon}{\sum_{k\in \text{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a} \right\}$;
(6) **return** z;

Algorithm 2: Max-min SINR.

the JLSPA problem, by looking for optimal and fair link transmissions strategies under SINR constraints, we aim to guarantee successful transmissions and incidentally reduce the schedule length. In fact, by guaranteeing MMF SINR, the number of potential concurrent links increases, which in turn implies shorter schedule length (see also Section 3). Hence, we present Algorithm 1 which solves the MMFSLT problem in Section 4.1. In order to consider the energy consumption in the network, we define another variant of the problem, called P_{energy}, in Section 4.1.

4.1. Solution Method for MMF SINR Link Computing. The MMFSLT problem aims to find a transmission power assignment scheme such that the concurrent transmissions in the network have a fair quality in terms of SINR. In order to achieve this goal, the first step is to maximize the minimum SINR value. However, it does not guarantee a MMF SINR for all the competitive links. More explicitly, the SINR value measured at all the receivers is not necessarily the same. Here, we go beyond this level and find the optimal max-min fair SINR for the set of competitive links. The basic steps for solving this problem are described in Algorithm 1.

Each step of Algorithm 1 is intended to find the z value that represents the max-min SINR value among all transmission links in a given set L. The respective link is identified, and the z value is allocated to its SINR value. Next, the algorithm removes the link from the set of links L and continues with the remaining ones. It stops iterating when z value achieves the β threshold or there is no link anymore in the set L. As a first step we define a problem called D_1. The z value is computed based on Algorithm 2. For the rest of the SINR values, we formulate and solve the problem D_k.

4.1.1. Formulating and Solving Problem D_1. The problem of maximizing the minimum value of SINR for a set of competitive links is modeled below:

$$\text{maximize} \quad \min_{j\in \text{RX}} \text{SINR}_j, \tag{5}$$

$$\text{s.t.:} \quad \text{SINR}_j \geq \alpha \quad \forall j \in \text{RX}, \tag{6}$$

$$\frac{P_i}{\omega_{ij}} \geq \text{RSSI}_0 \quad \forall i \in \text{TX}, \ (i, j) \in L, \tag{7}$$

$$P_{\min} \leq P_i \leq P_{\max} \quad \forall i \in \text{TX}, \tag{8}$$

where the first constraint (6) guarantees that the minimum SINR value is beyond the lower (α) threshold. The second one (7) ensures that the signal in the receiver is sufficiently high for being detected and processed. And the third emphasizes the fact that the node's power values should be in the interval $[P_{\min}, P_{\max}]$.

In the above formulation, we redefine the objective function. We denote by z the minimum value of SINR_j and add the respective constraints. Finally, the objective and the constraints of the problem D_1 are given as follows:

$$\text{maximize} \quad z, \tag{9}$$

$$\text{s.t.:} \quad \frac{P_i/\omega_{ij}}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a} \geq z \quad \forall \left(i,j\right) \in L, \tag{10}$$

$$\alpha \leq z, \tag{11}$$

$$\frac{P_i}{\omega_{ij}} \geq \mathrm{RSSI}_0 \quad \forall j \in \mathrm{RX}, \ \left(i,j\right) \in L, \tag{12}$$

$$P_{\min} \leq P_i \leq P_{\max} \quad \forall i \in \mathrm{TX}. \tag{13}$$

For this problem, the variables are given by P_i and z. We notice that the first constraint (10) is a nonlinear inequality. Nevertheless, we can easily handle this by fixing z to some lower bounds and keep increasing it appropriately until it reaches the optimal solution. We begin by initially setting $z = \alpha$ in the first constraint. Because α is a lower bound for z, this assumption leads to a feasible solution (if such a solution exists for the initial problem). By assuming z is a constant we obtain a LP model, the D_1' which is formulated as follows:

$$\text{maximize} \quad \epsilon, \tag{14}$$

$$\text{s.t.:} \quad \frac{P_i}{\omega_{ij}} - z \cdot \left(\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(\frac{P_k}{\omega_{kj}}\right) + N_a\right) \geq \epsilon \quad \left(i,j\right) \in L, \tag{15}$$

$$\alpha \leq z, \tag{16}$$

$$\frac{P_i}{\omega_{ij}} \geq \mathrm{RSSI}_0 \quad \forall j \in \mathrm{RX}, \ \left(i,j\right) \in L, \tag{17}$$

$$P_{\min} \leq P_i \leq P_{\max} \quad \forall i \in \mathrm{TX}. \tag{18}$$

In constraint (15) (of the problem D_1'), we have introduced the variable ϵ which will be helpful in increasing the z value (for the sake of simplicity we use the same notation z for both initial variable value and the current lower bound of z which is updated (increased) constantly through iterations). Thus, ϵ is such that for each sender i and receiver j

$$\frac{\epsilon}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a} \tag{19}$$

measures the gap between the SINR_j and the current z value. Hence, by increasing z according to (20) we ensure that there will be a feasible solution for the updated value of z:

$$\min_{j\in \mathrm{RX}} \frac{\epsilon}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a}. \tag{20}$$

Indeed, this is true since the last solution remains feasible for the updated z. The idea behind this is to gradually increase the z value until we reach its maximum value. Algorithm 2 describes these operations.

More precisely, the algorithm begins by solving the D_1' problem as defined above (with z set to α) and as a result we obtain the P_i values (or the power values assigned to sender nodes). The ratios between ϵ and interference at each receiver (see formula (20)) are used to obtain the minimum value that allows increasing the SINR values while guaranteeing a feasible solution. The process is finite and the algorithm will stop either when ϵ becomes practically 0 or when the inferior bound of SINR becomes larger than β. In the latter case we set $z = \beta$; otherwise we take the last z value. At this stage we cannot say too much on the theoretical complexity of the above algorithm. However, the methods perform quite well in practice and the process converges after a few steps. Notice last that, as suggested by an anonymous reviewer, the binary search could be a potential alternative method for computing the z value. Preliminary tests have not been concluding so we stuck to the epsilon method.

4.1.2. Formulating and Solving Problem D_k. Similarly to problem D_1, the problem D_k can be formulated as follows:

$$\text{maximize} \quad z, \tag{21}$$

$$\text{s.t.:} \quad \frac{P_i/\omega_{ij}}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a} \geq z \quad \forall \left(i,j\right) \in K, \tag{22}$$

$$\frac{P_i/\omega_{ij}}{\sum_{k\in \mathrm{TX}\setminus\{i\}} \left(P_k/\omega_{kj}\right) + N_a} \geq s_{(i,j)} \quad \forall \left(i,j\right) \in M \setminus K, \tag{23}$$

$$\frac{P_i}{\omega_{ij}} \geq \mathrm{RSSI}_0 \quad \forall j \in \mathrm{RX}, \ \left(i,j\right) \in M, \tag{24}$$

$$P_{\min} \leq P_i \leq P_{\max} \quad \forall i \in \mathrm{TX}. \tag{25}$$

In comparison with D_1, the above formulation differs in two points. First, constraints (10) give rise to two types of constraints, that is, (23) and (22), with respect to links with SINR already computed and the others. Second, constraint (11) is not useful any more as we have increasing values of z. Hence, any z solution to D_k is necessarily larger equal to precedent z and consequently to α. Problem D_k is solved in a similar way to problem D_1. More precisely, we use Algorithm 2 to solve D_k. To this end, we write an epsilon formulation, D_k', which is very similar to D_1'.

4.1.3. Identifying SINR Constrained Links. With respect to Algorithm 1, once Algorithm 2 has reached $\epsilon = 0$ and computed z, we need to find some links with SINR that cannot take higher value than z. Let us look in detail at problem D_1 and similar reasoning will hold for any D_k. Given problem (15)–(18) an easy way to find if a link is SINR constrained is to check if its dual coefficient of (15), let us say $\gamma_{(i,j)}$, is strictly positive. Indeed, using the complementary slackness property of duality theory, we have

$$\begin{aligned}&\left(\frac{P_i}{\omega_{ij}} - z \cdot \left(\sum_{k \in \mathrm{TX} \setminus \{i\}} \left(\frac{P_k}{\omega_{kj}}\right) + N_a\right) - \epsilon\right) \gamma_{(i,j)} \\ &\quad = \left(\frac{P_i}{\omega_{ij}} - z \cdot \left(\sum_{k \in \mathrm{TX} \setminus \{i\}} \left(\frac{P_k}{\omega_{kj}}\right) + N_a\right)\right) \gamma_{(i,j)} = 0\end{aligned} \tag{26}$$

and we can say that link (i, j) is SINR constrained if $\gamma_{(i,j)} > 0$. Furthermore, there is at least one strictly positive value, since from the dual formulation of the above problem we have $\sum_{(i,j)\in L} \gamma_{(i,j)} \geq 1$.

4.2. Correctness of MMF SINR Computation. Without loss of generality we assume that all z values computed during the algorithm are all less than β. The correctness of the proposed approach, that is, the optimality of the solution obtained by our iterative algorithm, is not obvious. Indeed, different transmission power assignments may satisfy all the constraints specified for the problem D_i at each step i of the algorithm. Furthermore, several links might possibly achieve the same SINR signal. Recall also that each step of Algorithm 2 yields at least one transmission link and the corresponding SINR value (z) with respect to the max-min fair SINR link transmissions. Once computed, this value is fixed and used as a constant for the remaining calculations. Questions naturally arise. How does this impact the upcoming z value and consequently the *quality* of the computed assignment? Which link should be chosen and what are the consequences for the desired power transmission assignment? In the following we will try to answer these questions and prove formally the optimality of the max-min fair assignment computed by Algorithm 1.

Theorem 1. *The power assignment solution obtained at the end of the kth step of Algorithm 1 is such that there is no other power assignment that would allow the SINR value of transmission links in L_k to be increased at the expense of other links with better SINR.*

Proof. We prove this by mathematical induction on the number of steps of Algorithm 1. Obviously the statement holds for the first step of the algorithm. Indeed, the way the set L_1 is defined makes the existence of some other solutions achieving better SINR for L_1 impossible. We will prove now that if the above property is true for any step $k-1$ then it is also true for the following step. Hence, by recurrence hypothesis we assume that there is no way of increasing any SINR in $L_{(k-1)}$ by decreasing SINR of the other links. At this stage we notice that any solution obtained at the end of step k is also a solution for all problems D_i, $i < k$, and all constraints (23) with respect to links in $M \setminus K$ are also satisfied at equality. Let us now consider some constrained link at step k; let us say (p, q). With respect to the formulation of the corresponding problem D'_k and bearing in mind that $\epsilon = 0$ and constraint (22) for link (p, q) is tight, it is clear that there can be no way of increasing the SINR value for link (p, q) simply by decreasing the SINR of some nonconstrained links (which are the only remaining links offering better SINR). Furthermore, this holds for all links in $M \setminus K$ as the current solution is as well a solution for all problems D_i, $i < k$, and the recurrence hypothesis applies. There is therefore no room to increase the SINR at this link and potential other links in L_k with the same SINR, which concludes the proof of the theorem. □

An immediate corollary of the above theorem is that the SINR values obtained at the end of Algorithm 1 give necessarily a max-min fair vector, since the result also holds for the solution obtained at the final step of the algorithm when all links are constrained.

4.3. Considering the Energy Consumption in the Network. The problem of MMFSLT computation aims to guarantee a "good" transmission medium for all concurrent links in the network. Hence, its focus falls upon the quality of links. In the WSN's context, energy is also considered as a relevant issue. Therefore, we can further process the results of Algorithm 1 in terms of economy of energy. By defining as objective the minimization of the sum of the nodes' power value, we formulate the problem P_{energy} as follows:

$$\text{minimize} \quad \sum_{i \in \mathrm{TX}} P_i, \tag{27}$$

$$\text{s.t.:} \quad \frac{P_i}{\omega_{ij}} - s_{(i,j)} \cdot \left(\sum_{k \in \mathrm{TX} \setminus \{i\}} \left(\frac{P_k}{\omega_{kj}}\right) + N_a\right) \geq 0 \quad \forall (i, j) \in M, \tag{28}$$

$$\frac{P_i}{\omega_{ij}} \geq \mathrm{RSSI}_0 \quad \forall j \in \mathrm{RX},\ (i, j) \in M, \tag{29}$$

$$P_{\min} \leq P_i \leq P_{\max} \quad \forall i \in \mathrm{TX}, \tag{30}$$

where the first constraint guarantees that the SINR value in each receiver is bigger than the $s_{(i,j)}$ threshold ($s_{(i,j)}$ in this case is considered a constant and it belongs to the S vector of the MMF values). The other constraints are identical with problem D_k. Nevertheless, notice that in practice there is no much space left for modifying power assignment when SINR values are determined for all active links. One way to deal with it stands in satisfying the SINR constraint only for the first level of MMF.

5. MMFSLT Problem for Time-Constant Transmission Power

In this section, we examine the case of MMFSLT problem with time-constant transmission powers. For a given network $G = (V, E)$, the transmission links E are allocated to different time slots with the same conditions as discussed in Section 3.2. The MMF algorithm (Algorithm 1) allocates the power to the transmitting nodes to guarantee MMF SINR at receiver for each given time slot. Hence, a node may have different transmission power depending on the time slot in which it transmits. Therefore, we formulate the following

Input: A set L of links located arbitrarily in the Euclidean plane;
Output: A feasible schedule S, the $\text{SINR}_{\text{threshold}}$;
while (*there are still links not assigned to a slot*) **do**
  $\Rightarrow$ take a new slot time;
  $\Rightarrow$ examine the non assigned links according to increasing distance to BS;
  $\Rightarrow$ assign link (j) to the current time slot if the function *test*(j) returns *True*;

test(j);
cond1: the link (j) and the other links in the current time slot have no receiver in common;
cond2: the sender node of link (j) and the receiver nodes of the other links in the current time slot are different;
cond3: assign a power for each sender node by solving problem D_1. The minimal value of the SINR evaluated for the set of links belonging to the current time slot, including link (j), is bigger than the $\text{SINR}_{\text{threshold}}$;
if (*cond1* & *cond2* & *cond3*) **then**
  return True;
else
  return False;

Algorithm 3: The principle of the algorithm for the JLSPA problem.

problem which seeks to find a unique transmission power for each node $\in V$ that is independent of the time slot and that ensures the SINR fairness. Let us denote by T the whole frame divided in time slots, RX_t the set of receiving nodes at slot t, TX_t the set of transmitting nodes at slot t, and M_t the set of activated links at slot t. An approach similar to the MMF algorithm can be given for the time-constant transmission power case at this stage: we report below the formulation of the counterpart of problem D_1:

$$\text{maximize} \quad \min \text{SINR}_{tj} \quad j \in \text{RX}_t,\ t \in T, \tag{31}$$

$$\text{s.t.:} \quad \frac{P_i}{\omega_{ij}} - \alpha \cdot \left(\sum_{k \in \text{TX}_t/\{i\}} \left(\frac{P_k}{\omega_{kj}} \right) + N_a \right) \geq 0, \quad (i, j) \in M_t, \quad t \in T, \tag{32}$$

$$\frac{P_i}{\omega_{ij}} \geq \text{RSSI}_0 \quad j \in \text{RX}_t,\ (i, j) \in M_t,\ t \in T, \tag{33}$$

$$P_{\min} \leq P_i \leq P_{\max} \quad \forall i \in \text{TX}. \tag{34}$$

6. The JLSPA Problem

Finally, we deal with the JLSPA problem. For solving the problem we need to perform the following tasks.

(1) Identify the activated links.

(2) Design a scheduling scheme.

(3) Assign the transmissions power.

6.1. Network Topology and Scheduling Algorithm. Here, we assume an uplink traffic in the network, according to a well-defined routing scheme (see Figure 2). Each sensor aggregates the data during the relaying (converge-cast with

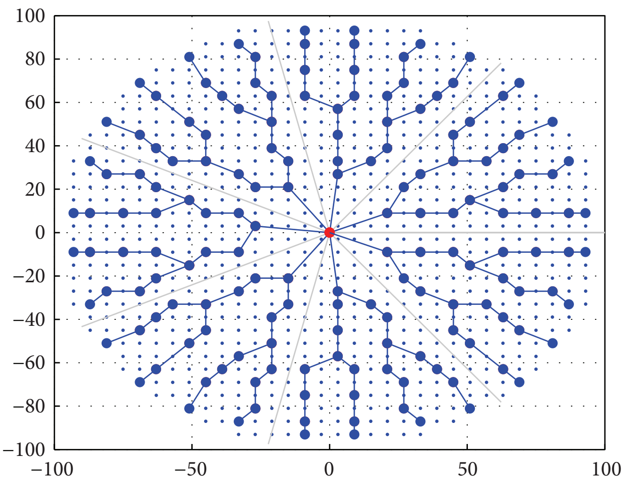


Figure 2: Routing tree: blue dots are the sensor nodes, and the red dot is the gateway.

data aggregation); therefore each transmission link may be activated only once. TDMA type of protocols following the same assumption is detailed in [28–30]. However, activating all the transmission links simultaneously may normally lead to unfeasible scenarios under SINR constraint.

Regarding the scheduling, it can be modeled as the one-shot scheduling problem in [16], assuming that the links weight will be equal to one unit. As this problem is shown to be $\mathcal{NP}$-hard, we propose a bottom-up approach described in Algorithm 3. This approach is a greedy heuristic which intends to put in a time slot the maximum number of links such that the SINR constraint is respected (where $\text{SINR}_{\text{threshold}}$ represents the α value; see (2)). The set of links L that need to be scheduled is given by solving the network configuration problem, as stated previously. In the set L

the links are sorted according to their respective coronas, meaning first we have the links of the corona closer to BS and next it will use the links of the second corona and so on. For each time slot, we try to put the links by beginning from those closer to the BS. The number of links that can be placed in a given time slot will be controlled by the three conditions given in lines 2–4 of the *test*() function. The algorithm proceeds by taking into consideration each link that has not been assigned to a time slot. If the function *test*() returns True, the link in consideration can be assigned to the current slot; otherwise the algorithm will check its validity for the next time slot.

The heart of this algorithm is the SINR computation (see condition 3) which corresponds to the third task, the power assignment strategy. For this we solve problem D_1; that is, we check the feasibility for a set of active links and compute the max-min SINR among them. Note that in the above approach we do not need to compute MMF SINR; nevertheless this can be done once the active links in a slot are determined. At this point one important question holds: how will the transmission power assignment strategy affect the schedule length? This question is answered in the following section.

6.2. Numerical Results for the JLSPA Problem. We apply two different power assignment strategies, (i) the linear power assignment and (ii) the power strategy for fair link transmissions (called MMF SINR strategy below), to the scheduling algorithm proposed previously. The linear power assignment strategy consists in assigning a power to each activated link, which is proportional to the link weight.

To compute the link weights we use the log-distance path loss model. This model is formally expressed according to

$$\omega_{ij}\left(d_{ij}\right) = \mathrm{PL}_0 + 10 \cdot \gamma \cdot \log_{10}\frac{d_{ij}}{d_0} + N\left(0,\sigma\right), \qquad (35)$$

where d_{ij} is the distance between transmitter (i) and receiver (j), d_0 a reference distance, PL_0 the power decay corresponding to d_0, γ the path loss exponent (rate at which signal decays), and $N(0,\sigma)$ a normal (or Gaussian) random variable with mean 0 and variance σ, reflecting the attenuation (in dB) caused by flat fading. The value of weight ω_{ij} is given in dBm. The link weights are computed according to the model given in (35), but for computation simplicity we have not considered the $N(0,\sigma)$ value. For the rest of parameters, the reader can refer to Table 1.

The MP parameters in Table 1 refer to the coefficients of MP presented in Section 4. For the α value we refer to the cochannel rejection ratio (CCRR) defined in the transceiver data sheets. From the empirical experiments provided in [31] we observed that the β value can be approximated at 10 dB. RSSI_0, $P_{\min}$, and $P_{\max}$ are extracted from the sensor (MICAz) data sheet.

Our algorithm is coded in C++ using the CPLEX 12.1 Library. The program is compiled with MSVC in a Windows environment, and all experiments were conducted on an AMD Opteron 2.60 GHz.

We have applied the linear and the fair power strategy to the scheduling problem and computed the schedule length for cases when the SINR threshold varies. These results are presented in Figure 3. The fair power assignment strategy improves the schedule length by at least 31% (fair power assignment strategy requires 18 time slots with respect to linear schedule length which needs 26 time slots for a SINR threshold equal to 1.9 units). As we can observe, the number of slots required to schedule all the links has the tendency to increase for bigger values of SINR thresholds.

Based on the above limited numerical results, it seems that the fair power assignment strategy can be helpful in reducing the schedule length. However, an extended numerical study would be necessary to confirm the findings while there is still place for further improving the scheduling heuristic.

Table 1: Simulation parameters.

Type	Parameter	Value
MP parameters	α	1.99 (3 dB)
	RSSI_0	−90 dBm
	$P_{\min}$	−25 dBm
	$P_{\max}$	0 dBm
	β	10 dB
	E_{rr}	10^{-7}
Channel parameters	γ	2
	Reference distance d_0	1 m
	Power PL_0	52.4
Radio parameters	Noise floor	−110 dBm
	White Gaussian noise N_w	4 dB
	$N(0,\sigma)$	0
Network topology instance	Network radius	100 m
	Internodes distance	6 m
	Number of activated links	161

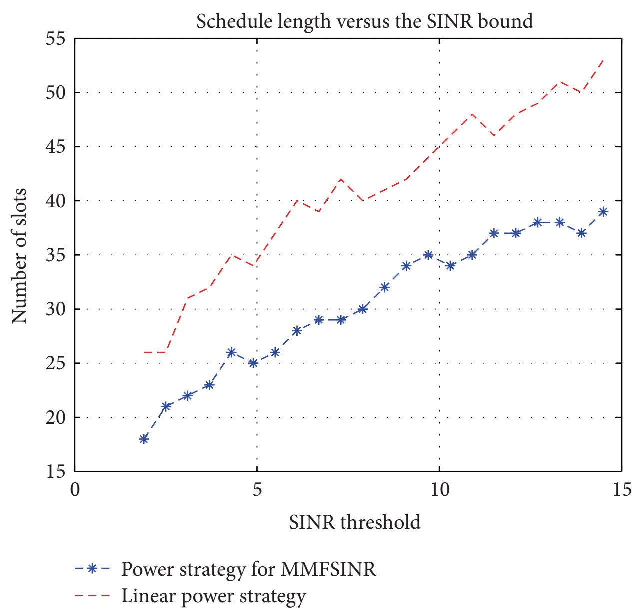


Figure 3: Time slots number versus SINR.

7. Future Work

We formulated and solved the max-min fair SINR link transmission problem which consists in allocating power transmission to nodes such that the SINR values of active links are MMF. This problem is solved optimally and, to the best of our knowledge, we are the first to propose an exact method. By implementing our method to JLSPA problem, we show that the schedule length may be significantly reduced. Moreover, the design of a cooperative approach provides a fair medium for the concurrent links and, therefore, the best possible scenario for having successful transmissions. We extended the problem also to the case when the transmission power of nodes is constant through the time frame. However, different problems may be interesting to be investigated as future work, such as

(i) improving the scheduling algorithm in order to have lower bounds for the JLSPA problem,

(ii) developing an adaptive transmission power control algorithm that operates in a distributed manner; the algorithm should be able to adapt its transmission power quickly to suit a rapidly changing radio environment.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

Part of this work has been performed in the AgentschapNL project Reconsurve-NL.

References

[1] D. Son, B. Krishnamachari, and J. Heidemann, "Experimental study of the effects of transmission power control and blacklisting in wireless sensor networks," in *Proceedings of the 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (SECON '04)*, pp. 289–298, October 2004.

[2] S. Lin, J. Zhang, G. Zhou, L. Gu, J. A. Stankovic, and T. He, "ATPC: adaptive transmission power control for wireless sensor networks," in *Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SenSys '06)*, pp. 223–236, November 2006.

[3] G. Hackmann, O. Chipara, and C. Lu, "Robust topology control for indoor wireless sensor networks," in *Proceedings of the 6th ACM Conference on Embedded Networked Sensor Systems (SenSys '08)*, pp. 57–70, November 2008.

[4] A. Gogu, D. Nace, A. Dilo, and N. Meratnia, "Review of optimization problems in wireless sensor networks," in *Telecommunication Networks, Current Status and Future Trends*, J. H. Ortiz, Ed., chapter 7, Intech, 2012.

[5] S. Chen, Y. Fang, and Y. Xia, "Lexicographic maxmin fairness for data collection in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 7, pp. 762–776, 2007.

[6] X. Wang, K. Kar, and J.-S. Pang, "Lexicographic max-min fair rate allocation in random access wireless networks," in *Proceedings of the 45th IEEE Conference on Decision and Control (CDC '06)*, pp. 1284–1290, December 2006.

[7] A. Sridharan and B. Krishnamachari, "Maximizing network utilization with max-min fairness in wireless sensor networks," in *Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '07)*, pp. 1–9, April 2007.

[8] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Proceedings of the Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '02)*, pp. 763–772, June 2002.

[9] D. Nace and M. Pióro, "Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.

[10] T. A. Al-Khdour and U. Baroudi, "An energy-efficient distributed schedulebased communication protocol for periodic wireless sensor networks," *Arabian Journal for Science and Engineering*, vol. 35, no. 2, pp. 153–168, 2010.

[11] S. C. Ergen and P. Varaiya, "TDMA scheduling algorithms for wireless sensor networks," *Wireless Networks*, vol. 16, no. 4, pp. 985–997, 2010.

[12] S. Gandham, M. Dawande, and R. Prakash, "Link scheduling in sensor networks: distributed edge coloring revisited," in *Proceedings of the 24th Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 4, pp. 2492–2501, March 2005.

[13] R. Kawano and T. Miyazaki, "Distributed data aggregation in multi-sink sensor networks using a graph coloring algorithm," in *Proceedings of the 22nd International Conference on Advanced Information Networking and Applications Workshops (AINA '08)*, pp. 934–940, March 2008.

[14] D. Chafekar, V. S. A. Kumar, M. V. Marathe, S. Parthasarathy, and A. Srinivasan, "Approximation algorithms for computing capacity of wireless networks with SINR constraints," in *Proceedings of the 27th Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '08)*, pp. 1840–1848, April 2008.

[15] T. Moscibroda and R. Wattenhofer, "The complexity of connectivity in wireless networks," in *Proceedings of the 25th Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '06)*, pp. 1–13, April 2006.

[16] O. Goussevskaia, Y. A. Oswald, and R. Wattenhofer, "Complexity in geometric SINR," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 100–109, September 2007.

[17] P. Santi, R. Maheshwari, G. Resta, S. Das, and D. M. Blough, "Wireless link scheduling under a graded SINR interference model," in *Proceedings of the 2nd ACM International Workshop on Foundations of Wireless Ad Hoc and Sensor Networking and Computing (FOWANC '09)*, pp. 3–12, May 2009.

[18] T. Kesselheim, "A constant-factor approximation for wireless capacity maximization with power control in the SINR model," in *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '11)*, pp. 1549–1559, 2011.

[19] Q. S. Hua, *Scheduling wireless links with SINR constraints [Ph.D. thesis]*, The University of Hong Kong, 2009.

[20] P. Gupta and P. R. Kumar, "Critical power for asymptotic connectivity in wireless networks," in *Stochastic Analysis, Control, Optimization and Applications*, pp. 547–566, Birkhäauser, Boston, Mass, USA, 1999.

[21] A. Fanghänel, T. Kesselheim, H. Räcke, and B. Vöcking, "Oblivious interference scheduling," in *Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC '09)*, pp. 220–229, 2009.

[22] A. Gogu, S. Chatterjea, D. Nace, and A. Dilo, "The problem of joint scheduling and power assignment in wireless sensor networks," in *Proceedings of the 27th IEEE International Conference on Advanced Information Networking and Applications (AINA '13)*, pp. 348–355, March 2013.

[23] P. Björklund, P. Värbrand, and D. Yuan, "Resource optimization of spatial TDMA in ad hoc radio networks: a column generation approach," in *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies (INFOCOM '03)*, vol. 2, pp. 818–824, April 2003.

[24] S. A. Borbash and A. Ephremides, "Wireless link scheduling with power control and SINR constraints," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 5106–5111, 2006.

[25] B. Katz, M. Völker, and D. Wagner, "Energy efficient scheduling with power control for wireless networks," in *Proceedings of the 8th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '10)*, pp. 160–169, June 2010.

[26] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1029–1046, 2009.

[27] A. Sridharan and B. Krishnamachari, "Max-min fair collision-free scheduling for wireless sensor networks," in *Proceedings of the 23rd IEEE International Performance, Computing, and Communications Conference (IPCCC '04)*, pp. 585–590, April 2004.

[28] S. Chatterjea, L. F. W. van Hoesel, and P. J. M. Havinga, "AI-LMAC: an adaptive, information-centric and lightweight MAC protocol for wireless sensor networks," in *Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference (ISSNIP '04)*, pp. 381–388, Melbourne, Australia, December 2004.

[29] L. van Hoesel and P. Havinga, "A lightweight medium access protocol (LMAC) for wireless sensor networks," in *Proceedings of the 1st International Workshop on Networked Sensing Systems (INSS '04)*, Tokyo, Japan, 2004.

[30] K. Langendoen, "Medium access control in wireless sensor networks," in *Medium Access Control in Wireless Networks Volume II: Practice and Standards*, chapter 20, pp. 535–560, Nova Science, 2008.

[31] M. Z. Zamalloa and B. Krishnamachari, "An analysis of unreliability and asymmetry in low-power wireless links," *ACM Transactions on Sensor Networks*, vol. 3, no. 2, Article ID 1240227, 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 562373, 12 pages
http://dx.doi.org/10.1155/2014/562373

Research Article

An Approximation Algorithm for the Facility Location Problem with Lexicographic Minimax Objective

Ľuboš Buzna, Michal Koháni, and Jaroslav Janáček

Department of Transportation Networks, Faculty of Management Science and Informatics, University of Žilina, Univerzitná 8215/1, 01026 Žilina, Slovakia

Correspondence should be addressed to Ľuboš Buzna; buzna@frdsa.uniza.sk

Received 30 January 2014; Accepted 24 April 2014; Published 17 June 2014

Academic Editor: Wlodzimierz Ogryczak

We present a new approximation algorithm to the discrete facility location problem providing solutions that are close to the lexicographic minimax optimum. The lexicographic minimax optimum is a concept that allows to find equitable location of facilities serving a large number of customers. The algorithm is independent of general purpose solvers and instead uses algorithms originally designed to solve the p-median problem. By numerical experiments, we demonstrate that our algorithm allows increasing the size of solvable problems and provides high-quality solutions. The algorithm found an optimal solution for all tested instances where we could compare the results with the exact algorithm.

1. Introduction

Our study is motivated by problems faced by public authorities when locating facilities, such as schools, branch offices, or libraries and ambulance, police, or fire stations. To find an efficient spatial design of these systems, various types of location and allocation models can be used [1–4]. Typically, these problems are seen as an example of the resource allocation problem with a central planner. Thus, the costs of the system construction and its maintenance are typically "shared" by everybody, though not all contributors to the system (tax payers) are enjoying the same access to services. When pursuing an economically efficient design, some customers may end up being located close to the located service centres, whereas others are placed far away. Self-interested customers could understand such situation as unfair.

It is clear that in a real-world situation it is impossible to reach completely equal access to services for all customers, however, considering fairness criteria can help distribute the accessibility of services among customers more evenly. The number of existing problems related to the fair division of scarce resources is overwhelming. We restrict our short overview to few examples only. A recent comprehensive overview of models, algorithms, and applications is available in [5].

In manufacturing, the omnipresent problem is how to allocate limited resources among many competing processes. Solving methods for problems with the knapsack inequality type of constraints, including multiple resource constraints, multiperiod problems, and problems with substitutable resources, are discussed in [6]. In engineering, one of the most profound applications of fairness is sharing of capacities in communication networks [7–9]. The method for solving the maximum flows [10] often generates unfair flows in the sense of how the flows are distributed to sink nodes. Using the principle of max-min fairness [11], the sink- (source-) optimal flow was defined as a flow which lexicographically maximizes the flow vector from the point of view of sink (source) nodes, and an efficient algorithm to find such flows was proposed in [12, 13]. The same fairness scheme was studied in the context of various routing mechanisms. An overview of basic problems in communication networks, associated with the applications of the max-min fairness to the flow rate allocation, routing, and load balancing, is given in [14]. For example, fixed paths were studied in regular networks finding analytical expressions describing the network throughput [15]. For multicommodity flows, LP formulation of the problem was used to represent the flexible routing [16]. When mixed or integer decision problems are to be solved, the basic sequential procedure is not applicable. The ordered outcomes approach and the ordered values approach [17] allow to overcome this difficulty. Application of these two approaches to the bandwidth allocation problem was demonstrated in [18].

One important element appearing in studies focusing on fairness is the efficiency-fairness tradeoff. A well-known limitation of the lexicographically maximal flow is the relatively large reduction in the network throughput. A possible solution can be found in the optimization of flows with respect to the efficiency, provided that some level of fairness is guaranteed. An example of such approach is (α, β) fairness introduced in [19]. A general approach, combining optimization of the minisum and the minimax criteria with the tunable size of the applicability area for the minimax criterion, was proposed in [20]. To demonstrate the applicability of the proposed approach, the authors analysed the problem of how to distribute limited resources among patients to pay for the costly medical treatments. Alternatively, one can use different fairness schemes. One such scheme known as the proportional fairness was generalized and presented as an optimization problem [21]. The efficiency of the max-min fairness and proportional fairness was studied and analytically evaluated for a general set-up, using a simple measure [22], finding analytical expressions describing bounds for a gap between fair and efficient solutions. Using the generalized objective function which encompasses both these schemes as special cases, the same authors studied the efficiency-fairness trade-off, proposing several managerial prescriptions for the selection of the objective [23].

The first attempts to consider equity considerations when solving location problems on networks date back to the influential paper [24]. In this paper, the problem of finding the minimum number of locations was addressed, considering that no customer is farther from an existing location than a given distance. Since then, the inclusion of equity in location models has been a recurring topic. The equity is usually quantified by an equity measure. The taxonomy of equity measures proposed specifically for location problems was given in [25]. Here the authors decided to organize the equity measures around three dimensions: the choice of reference distribution which represents the desired goal, the distance metric which determines the way of how to assess the distance to the desired goal, and the scaling function used to take into account different importance of customers (e.g., by considering population, land area, demand, or income). In addition, the authors summarize from the literature useful criteria which should be taken into account when selecting an equity measure. Recently, an application of the equity measure to the equitable facility location problem in a plane was described in [26]. The authors analysed properties of the Gini coefficient and proposed an algorithm that finds the optimal location of a facility in a bounded area.

The requirement of equitable distribution comes often combined with other objectives. Noteworthy is the close relation between fairness criteria and multicriterial optimization. Typically, the equity is either formally represented by one out of several criteria or the interest of each individual is represented as a single objective function [27]. For the survey of multiple criteria facility location problems, including those considering the equity, see the recent paper by Farahani et al. [28].

In the literature we do not find many attempts to suggest algorithms for solving facility location problems considering the lexicographic minimax objective. Specialized algorithms, considering minimax and lexicographic minimax optimization, were proposed when locating single facility in a plane [29]. Another approach, applicable to planar problems simultaneously optimizing the equitable distribution of distances by minimizing the radius of the serviced area and ensuring equitable distribution of loads, is described in [30]. The problem of how to locate facilities with equitable loads using the minimax criterion was also studied on networks [31]. The work [32] is closely related to the issue of lexicographic minimax optimization on networks. The authors noticed the possibility of reducing the classical minimax to minisum for 0-1 programming problems by transforming the coefficients in the objective function using a power function [32]. As pointed out in [33], the concept of power functions can be also extended to the lexicographic minimax solution concept. However, in general, high powers may be necessary to generate large enough differences between distances. When solving practical problems, large differences between distance coefficients may cause serious computational problems. To overcome these difficulties, the ordered outcomes approach and the ordered values approach were proposed in [17]. The latter approach was found as more efficient, but the size of tested instances was rather small. A very convenient technique for interactive analysis, where facilities are located with respect to the objective function taking into account lexicographic minimax combined with the minisum term, was proposed in [34]. The approach is based on the reference distribution method which can be controlled by manipulating few parameters and allows to take into account aspiration values of assigned distances defined by the user.

The lexicographic minimax optimization problem can be converted to a problem, where the kth term in the objective function is the number of occurrences of the kth worst possible unique outcome. The optimal solution is then found by minimizing the first term followed by the minimization of the second term without worsening the first term and so on [5]. The approach to the facility location problem based on this concept was developed in [33]. The initial computational experience with this approach shows that the lexicographic minimax approach, in comparison to the standard minimax, selects the locations characterized by remarkably smaller mean distance and absolute difference. Although the algorithm requires to run a large number of stages for the general distance matrix, it performs very well for the cases when the matrix contains only few distinct distance values. This makes the algorithm a very good choice when searching for an approximate solution or when only rough estimates of distances are considered.

Each of the above mentioned approaches to the lexicographic minimax optimization [17, 33, 34] results in a specific form of the mathematical model that is supposed to be solved by a general purpose solver. Our initial experience with the algorithm [33], implemented on the state of the art solver XPRESS [35], indicated that we are able to solve problems up to 900 customers and 900 candidate facility locations, while restricting the distances to integers and measuring them in kilometres. This limitation might be too tight for some real-world applications and therefore it is of interest to elaborate algorithms which can provide high-quality solutions to larger problems. Building on the same basic concept of the unique classes of distances [33], we propose an approximation algorithm providing high-quality solutions for large instances of solved problems. We use the resulting algorithm to perform an extensive study using the well-known benchmarks [36–39], and we enriched them by new benchmarks derived from the real-world road network data [40, 41]. Our main contribution is the set of rules which allow us (i) to take into account the multiplicities assigned to different customers; (ii) to detect whether for a given distance active customers can reach higher, equal, or smaller distance to the closest located facility; and (iii) to use methods customized for solving the p-median problem. Customized methods can handle larger problems than up-to-date general purpose integer programming solvers. Therefore, the applicability of the algorithm is enhanced, especially, when it is used to solve large instances of real-world problems.

The remainder of the paper is organized as follows: the lexicographic minimax approach to the facility location problem is briefly described in Section 2.1. Section 2.2 presents our algorithm. In Section 3 the benefits of the algorithm are demonstrated on the set of benchmarks derived from the real-world networks. To conclude, we summarize our findings in Section 4.

2. Materials and Methods

When solving a facility location problem, the goal is to find suitable positions of facilities that provide services to the set of customers distributed in a serviced area. Along the years, many variants of the facility location problem have been elaborated. As an archetypical example, we describe the weighted p-median problem.

2.1. The Equitable Facility Location Problem. We consider the set I of potential locations for facilities and the set of locations J representing aggregate customers. Each aggregate customer $j \in J$ is characterized by a unique geographical position and we associate an integer weight (multiplicity) b_j with each position. The weight b_j represents the number of individual customers situated in the location j. We denote the set of all individual customers by $\bar{J}$. In order to map individual customers to aggregate customers we define the function $j(k)$ for $k \in \bar{J}$, returning the element $j \in J$ if and only if the individual customer k is situated in the location j (see Figure 1).

The decisions to be made can be represented by a set of binary variables. The variable y_i equals 1 if the location $i \in I$ is used as a facility location and equals 0 otherwise. Allocation

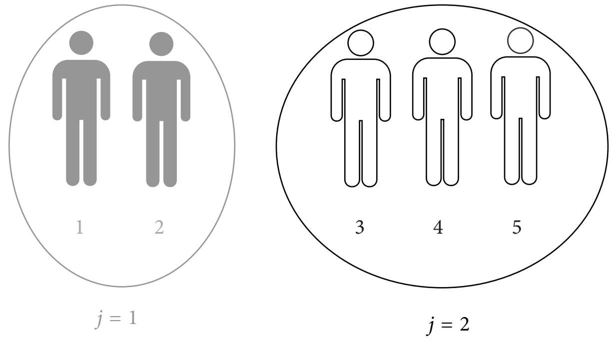


Figure 1: Schematic illustrating the definition of customers. The set of aggregate customers J is composed of two elements, that is, $J = \{1, 2\}$, and the set of individual customers is $\bar{J} = \{1, 2, 3, 4, 5\}$. The aggregate customer 1 includes individual customers 1 and 2, and the aggregate customer 2 stands for individual customers 3, 4, and 5. Thus, in this case the function $j(k)$, mapping the individual customer $k \in \bar{J}$ to the aggregate customer, returns these values $j(1) = 1$, $j(2) = 1$, $j(3) = 2$, $j(4) = 2$, and $j(5) = 2$, and the weights (multiplicities) assigned to aggregate customers are $b_1 = 2$ and $b_2 = 3$.

decisions are modelled by variables x_{ij} for $i \in I$ and $j \in J$, whereas $x_{ij} = 1$ if location i is serving the customer j and $x_{ij} = 0$ otherwise. In order to obtain a feasible solution, the decision variables have to satisfy the following set of constraints:

$$\sum_{i \in I} y_i = p, \tag{1}$$

$$\sum_{i \in I} x_{ij} = 1 \quad \forall j \in J, \tag{2}$$

$$x_{ij} \le y_i \quad \forall i \in I,\ j \in J, \tag{3}$$

$$x_{ij} \in \{0, 1\} \quad \forall i \in I,\ j \in J, \tag{4}$$

$$y_i \in \{0, 1\} \quad \forall i \in I, \tag{5}$$

where (1) specifies that the number of located facilities equals p. The constraints (2) make sure that each customer is assigned to exactly one facility, and the constraints (3) allow assigning a customer only to the located facilities. Following [33], we denote the set of all feasible location patterns, which satisfy the constraints (1)–(5), by the symbol Q. For each customer $j \in J$ we define the performance function $f_j(\mathbf{x})$ [27] where $\mathbf{x}$ is a matrix representing the assignment decisions. This function estimates how customer j values the effect of located facilities. There is a strong stream of literature in location science studying measures that can be used to describe preferences of customers [42]. Typically, a measure called utility (disutility), u_{ij}, is defined to characterize preference of the customer j for the candidate location i. Utility is often defined as a function of the distance d_{ij} from the customer j to the facility location i. The performance function $f_j(\mathbf{x})$ is then a composite of utilities u_{ij} representing a model of the customer's choice behaviour [43]. The two most frequently used choice behavioural models are the minimum function, when the demand of a customer is assigned to one located facility only (e.g., the closest one), and some kind of gravity model when the customer's demand is distributed among located facilities following a given mathematical prescription. Although the presented approach is able to handle the situation when u_{ij} is an arbitrary function of the distance d_{ij}, to simplify the text we choose $u_{ij} = d_{ij}$. The main focus of this paper is to extend the size of problems solvable by lexicographic approach, and, therefore, for simplicity reasons, we define the performance function that represents the distance from the customer to the closest located facility. Consequently, individual customers will get the same value of the performance function if they correspond to the same aggregate customer. Using constraints (2), the performance function for the individual customer $k \in \overline{J}$ can be formulated as

$$f_k(\mathbf{x}) = f_j(\mathbf{x}) = \sum_{i \in I} d_{i,j} x_{i,j}, \tag{6}$$

where $j = j(k)$ and d_{ij} is the distance from the aggregate customer $j \in J$ to the facility location $i \in I$.

The system optimum [22], frequently referred to as the minisum optimum or the utilitarian solution, corresponding to the p-median problem [1] is obtained when we minimize the expression (7) subject to $(\mathbf{x}, \mathbf{y}) \in Q$:

$$S(\mathbf{x}) = \sum_{j \in J} b_j f_j(\mathbf{x}) = \sum_{k \in \overline{J}} f_k(\mathbf{x}). \tag{7}$$

The corresponding optimal solutions we denote

$$\left(\mathbf{x}^{\mathrm{SYS}}, \mathbf{y}^{\mathrm{SYS}}\right) = \arg\min \{S(\mathbf{x}) \mid (\mathbf{x}, \mathbf{y}) \in Q\}. \tag{8}$$

The standard definition of the lexicographic minimax optimum [33] can be adjusted to the weighted problem, where the weights are representing the multiplicities of customers, as follows. We enlarge the set of aggregate customers J to the set of individual customers $\overline{J}$; that is, each aggregate customer $j \in J$ is replaced by b_j individual customers situated in the same location. After this adjustment, we introduce the map $\Theta : R^{|\overline{J}|} \to R^{|\overline{J}|}$ which orders the values $f_k(\mathbf{x})$ for $k \in \overline{J}$ in a nonincreasing order. Thus, more formally,

$$\begin{aligned} &\Theta\left(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_{|\overline{J}|}(\mathbf{x})\right) \\ &\quad = \left(\overline{f}_1(\mathbf{x}), \overline{f}_2(\mathbf{x}), \ldots, \overline{f}_{|\overline{J}|}(\mathbf{x})\right), \end{aligned} \tag{9}$$

if and only if there is the permutation τ such that $\overline{f}_k(\mathbf{x}) = f_{\tau(k)}(\mathbf{x})$ for all $k \in \overline{J}$ where $\overline{f}_1(\mathbf{x}) \geq \overline{f}_2(\mathbf{x}) \geq \cdots \geq \overline{f}_{|\overline{J}|}(\mathbf{x})$. Let us denote

$$\begin{aligned} \mathbf{v} &= \left(\overline{f}_1(\mathbf{x}^1), \overline{f}_2(\mathbf{x}^1), \ldots, \overline{f}_{|\overline{J}|}(\mathbf{x}^1)\right), \\ \mathbf{u} &= \left(\overline{f}_1(\mathbf{x}^2), \overline{f}_2(\mathbf{x}^2), \ldots, \overline{f}_{|\overline{J}|}(\mathbf{x}^2)\right), \end{aligned} \tag{10}$$

where $(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2) \in Q$. We define the strong lexicographic inequality $<_{\mathrm{LEX}}$ by $\mathbf{v} <_{\mathrm{LEX}} \mathbf{u}$ if and only if there is an index $m \leq |\overline{J}|$ such that $v_n = u_n$ for all $n < m$ and $v_m < u_m$. The weak lexicographic inequality $\leq_{\mathrm{LEX}}$ is then defined as $\mathbf{v} \leq_{\mathrm{LEX}} \mathbf{u}$, if and only if either $\mathbf{v} = \mathbf{u}$ or $\mathbf{v} <_{\mathrm{LEX}} \mathbf{u}$. Now we can define solution $(\mathbf{x}^{\mathrm{LEX}}, \mathbf{y}^{\mathrm{LEX}})$ to be the lexicographic minimax optimum if

$$\begin{aligned} &\Theta\left(f_1(\mathbf{x}^{\mathrm{LEX}}), f_2(\mathbf{x}^{\mathrm{LEX}}), \ldots, f_{|\overline{J}|}(\mathbf{x}^{\mathrm{LEX}})\right) \\ &\quad \leq_{\mathrm{LEX}} \Theta\left(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_{|\overline{J}|}(\mathbf{x})\right), \end{aligned} \tag{11}$$

for all $(\mathbf{x}, \mathbf{y}) \in Q$.

We defined the lexicographic ordering and the lexicographic optimum using the set of individual customers. When we translate this definition to the set of aggregate customers, we can interpret the lexicographic optimization as two subsequent goals. Primarily, we are aiming to assign to facilities those customers whose performance function cannot be lowered any further; that is, the maximal distance from the relevant customers to the closest facility is minimized. Secondarily, we minimize the number of relevant customers which are assigned to a facility at the considered stage. Please note that the number of individual customers, having assigned a distance value, is given by the sum of b_j values corresponding to those aggregate customers that are decided upon. In the next section we present the detailed description of the algorithm.

2.2. *The Approximation Algorithm to the Facility Location Problem with the Lexicographic Minimax Objective.*

Similarly to the algorithm [33], our algorithm subsequently solves optimization problems corresponding to the distance values in stages. We order the set of all feasible distance values $d_{i,j}$ into the descending sequence of unique distance values D_k, for $k = 1, \ldots, k_{\max}$. At each stage $k > 1$ we consider a partitioning of the set J into the system of subsets $\{J_1, \ldots, J_{k-1}, C_k\}$, where C_k is a set of active customers. We aim to identify the minimal subset of customers $J_k \subseteq C_k$, whose distance from the closest facility location equals the value D_k. We define the minimal subset as the set with the minimum number of individual customers, that is, the set where the sum of multiplicities $\sum_{j \in J_k} b_j$ is the smallest. For a given value of D_k, we find the minimal set J_k by solving the problem P_k:

$$\begin{aligned} &\text{Minimize} \quad g^k(\mathbf{x}) = \sum_{i \in I} \sum_{j \in J} r_{ij}^k x_{ij} \\ &\text{Subject to} \quad (\mathbf{x}, \mathbf{y}) \in Q, \end{aligned} \tag{12}$$

where r_{ij}^k are the costs defined for $j \in C_k$ and $i \in I$ in the following way:

$$r_{ij}^k = \begin{cases} 0, & \text{if } d_{ij} < D_k, \\ b_j, & \text{if } d_{ij} = D_k, \\ \left(1 + \sum_{u \in C_k} b_u\right), & \text{if } d_{ij} > D_k, \end{cases} \tag{13}$$

and for $j \in J_l$ where $l = 1, \ldots, k-1$ and $i \in I$ according to the following prescription:

$$r_{ij}^k = \begin{cases} 0, & \text{if } d_{ij} \leq D_l, \\ \left(1 + \sum_{u \in C_k} b_u\right), & \text{otherwise.} \end{cases} \quad (14)$$

This setting of coefficients r_{ij}^k allows us to effectively distinguish three important situations, which can be directly used in the construction of the algorithm. Knowing the optimal solution $(\mathbf{x^k}, \mathbf{y^k})$ of the problem P_k, the following implications denoted as cases (a), (b), and (c) can be derived.

(a) If $g^k(\mathbf{x^k}) = 0$, then each customer $j \in C_k$ can be assigned to a facility whose distance from j is less than D_k, and each customer $j \in J_l$ for $l = 1, \ldots, k-1$ can be assigned to a facility whose distance from j is less than or equal to D_l.

(b) If $0 < g^k(\mathbf{x^k}) < 1 + \sum_{u \in C_k} b_u$, then each customer $j \in C_k$ can be assigned to a facility whose distance from j is less than or equal to D_k, and each customer $j \in J_l$ for $l = 1, \ldots, k-1$ can be assigned to a facility whose distance from j is at most D_l. The minimal subset of customers $J_k \subseteq C_k$ whose distance from the closest facility locations equals the value D_k can be defined as $\{j \in C_k \mid \sum_{i \in I} r_{ij}^k x_{ij}^k = b_j\}$.

(c) If $g^k(\mathbf{x^k}) > \sum_{u \in C_k} b_u$, then there exists either the customer $j \in C_k$, which is farther from the allocated facility than D_k or a customer in the subset J_l that is farther from the allocated facility than D_l. Thus, this case indicates nonexistence of a solution $(\mathbf{x}, \mathbf{y})$ to the problem P_k, for which $\sum_{i \in I} d_{ij} x_{ij} \leq D_l$ for $j \in J_l$, where $l = 1, \ldots, k$.

We formulate the following algorithm, where we identify the customers whose distance from the closest facility location cannot be shorter than D_k, by embedding the problem P_k:

Algorithm A-LEX

Step 0. Initialize $k = 1$ and $C_1 = J$.

Step 1. Solve the problem P_k and denote the optimal solution by $(\mathbf{x^k}, \mathbf{y^k})$.

Step 2. If $g^k(\mathbf{x^k}) = 0$, set $C_{k+1} = C_k$ and go to Step 4; otherwise if $(0 < g^k(\mathbf{x^k}) < 1 + \sum_{u \in C_k} b_u)$ go to Step 3.

Step 3. Set $J_k = \{j \in C_k \mid \sum_{i \in I} r_{ij}^k x_{ij}^k = b_j\}$; $C_{k+1} = C_k - J_k$.

Step 4. If $C_{k+1} = \emptyset$, then terminate and return $(\mathbf{x^k}, \mathbf{y^k})$ as the solution; otherwise set $k = k + 1$ and continue with Step 1.

Correctness and finiteness of the algorithm A-LEX are justified by the following propositions.

Proposition 1. *The optimal solution of the problem P_k cannot satisfy the inequality $g^k(\mathbf{x^k}) > \sum_{u \in C_k} b_u$.*

Proof. For $k = 1$, the sum of all coefficients r_{ij}^1 does not exceed the value $\sum_{u \in C_1} b_u$ and thus $g^1(\mathbf{x^1}) \leq \sum_{u \in C_1} b_u$.

For $k > 1$, let us assume that the solution $(\mathbf{x^{k-1}}, \mathbf{y^{k-1}})$ of the problem P_{k-1} complies either with the case (a) or with the case (b) and $C_k \neq \emptyset$. The solution $(\mathbf{x^{k-1}}, \mathbf{y^{k-1}})$ assigns all customers from the set $C_k = C_{k-1} - J_{k-1}$ to facilities that are distant by at most D_k.

For $l = 1, \ldots, k-1$, the customer $j \in J_l$ is assigned to facility at the distance D_l and therefore the following inequality $g^k(\mathbf{x^{k-1}}) < 1 + \sum_{u \in C_k} b_u$ is valid. Solution $(\mathbf{x^k}, \mathbf{y^k})$, as a minimizer of the problem P_k, must fulfil inequality $g^k(\mathbf{x^k}) \leq g^k(\mathbf{x^{k-1}})$ and consequently $g^k(\mathbf{x^k}) \leq \sum_{u \in C_k} b_u$. □

To assure the consistent termination of the algorithm A-LEX, the set C_k must be emptied for $k \leq k_{\max} + 1$.

Proposition 2. *If $C_{k_{\max}} \neq \emptyset$, then $C_{k_{\max}+1} = \emptyset$.*

Proof. If $C_{k_{\max}} \neq \emptyset$, then the solution $(\mathbf{x^{k_{\max}}}, \mathbf{y^{k_{\max}}})$ assigns all customers in the set $C_{k_{\max}}$ to facilities that are at the distance $D_{k_{\max}}$. As $D_{k_{\max}}$ is the minimal distance value, $J_{k_{\max}} = C_{k_{\max}}$ and thus $C_{k_{\max}+1} = \emptyset$. □

To investigate under what circumstances the algorithm A-LEX provides the optimal solution, we start by noting that each feasible solution in the set Q is associated with a sequence of sets $[J_1, J_2, \ldots, J_{k_{\max}}]$ and with a vector $[B_1, B_2, \ldots, B_{k_{\max}}]$. The distance between customers in the set J_k and the assigned facility is exactly D_k. The component B_k is a number defined as $B_k = \sum_{j \in J_k} b_j$. If a set J_k is empty, then the associated value B_k is zero. The lexicographically minimal solution in the set Q corresponds to the lexicographically minimal vector $[B_1, B_2, \ldots, B_{k_{\max}}]$ [5].

Proposition 3. *If algorithm A-LEX does not find the lexicographically minimal solution, then there must exist a problem P_k having at least two optimal solutions $(\mathbf{x^A}, \mathbf{y^A})$ and $(\mathbf{x}^*, \mathbf{y}^*)$ associated with two different sets J_k^A and J_k^*, respectively.*

Proof. Let us consider that the algorithm A-LEX found solution $(\mathbf{x^A}, \mathbf{y^A})$, which is not lexicographically optimal; then the optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$ induces vector $[B_1^*, B_2^*, \ldots, B_{k_{\max}}^*]$ that is lexicographically smaller than the vector $[B_1^A, B_2^A, \ldots, B_{k_{\max}}^A]$. Let us denote by k_0 the smallest subscript, for which the inequality $B_{k_0}^* < B_{k_0}^A$ holds.

Assuming that no pair $[J_k^*, J_k^A]$ consists of different sets for $k = 1, \ldots, k_0 - 1$, then we obtain $C_{k_0}^A = J - \cup_{k=1}^{k_0-1} J_k^A = J - \cup_{k=1}^{k_0-1} J_k^* = C_{k_0}^*$.

As A-LEX solves P_{k_0} to optimality, the assumption of identity for J_k^A and J_k^* for $k = 1, \ldots, k_0 - 1$ contradicts the inequality $B_{k_0}^* < B_{k_0}^A$. □

Corollary 4. *If there is only one optimal solution of the problem P_k concerning variables $\mathbf{x^k}$ at each stage of the algorithm, then the solution provided by the algorithm A-LEX is lexicographically optimal.*

It is important to note that the algorithm A-LEX yields an optimal solution with only minor exceptions concerning the ambiguous determination of the set J_k. A necessary condition for obtaining an approximate solution is that at the stage k there exist at least two optimal solutions of the problem P_k. Thus, there exist at least two different sets of customers J_k^* and J_k^A such that $\sum_{j\in J_k^*} b_j = \sum_{j\in J_k^A} b_j$, and each set of customers is assigned the distance D_k. Our algorithm is able to identify only one of these sets and it treats all such sets as being equivalent. An approximation error may arise when the algorithm chooses the set, which in the future steps leads to the distribution of distances that is lexicographically less efficient than the distribution that would be reached, when continuing with the alternative set of customers. The size of the approximation error depends on how likely the multiple solutions of the problem P_k are and whether multiple solutions have the tendency to appear in the early stages of the algorithm or close to the end. Thus, the resulting error depends on the distribution of b_j and d_{ij} values. This drawback is compensated by several advantages, which enable speeding up significantly the algorithm and thus enable solving larger problems. As the sequence of problems P_k keeps the same size and structure of constraints, we can easily replace the general purpose IP solver by an algorithm customized for the p-median problem.

When processing D_k values in the descending order one by one, we may observe that sets J_k are empty for a range of k values, especially at the beginning of the solving process (for large D_k values). Such "empty" iterations do not have any impact on the resulting solutions and could be skipped. As the processing of the problem P_k allows us to identify situations in which there are customers that require to be assigned larger distance value than D_k, values D_k do not have to be processed sequentially. We can use various searching schemes (e.g., bisection method) to find the next value D_k where $J_k \neq \emptyset$. This can help reduce the number of times the problem P_k is solved and thus reduce the overall computational time.

To initialize the algorithm, we do not need to consider the entire sequence of values D_k for $k = 1, \ldots, k_{\max}$. Instead, we can start the solving process by finding the optimum corresponding to the objective function (7). Then it is enough to initialize the variable k in Step 0 to the value m that is determined by the equality $D_m = \max\{d_{ij} \mid x_{ij}^{\mathrm{SYS}} = 1\}$. This typically allows to skip many processing steps when no customer is assigned a distance value. Another useful operation (for $k > 1$) is to check whether $g^k(\mathbf{x}^{\mathbf{k-1}}) = 0$ in Step 1 before the problem P_k is solved. If that is the case, we can use $\mathbf{x}^{\mathbf{k-1}}$ as the optimal solution of the problem P_k [33].

3. Results and Discussion

To study the efficiency of the algorithm A-LEX, we performed a computational study. Our two main goals were to evaluate the quality of the solutions by comparing them to the exact algorithm [33] (algorithm O-LEX hereafter) and to test the limits of the algorithm regarding the size of the solvable problems. The algorithm O-LEX adds new constraints to the solved problem. Consequently, the use of algorithms customized for the p-median problem is impossible. Therefore we used a general purpose IP solver to implement it. To be able to compare the algorithm A-LEX with the algorithm O-LEX, we implemented them in the XPRESS-Mosel language (version 3.4.0) and we executed them using the XPRESS-Optimizer (version 23.01.05) [35]. To keep both algorithms comparable, we did not use any searching scheme to process values D_k in the algorithm A-LEX, and thus we processed all values D_k sequentially. To explore the properties of the algorithm A-LEX beyond the limits of the general purpose integer solvers, we implemented the algorithm A-LEX in the Microsoft Visual C++ 2010. To solve the problem (12) we used the exact algorithm ZEBRA, the state of the art solver for the p-median problem [44]. The implementation of the algorithm ZEBRA is publicly available on the author's web page [45]. To distinguish both versions of the algorithm A-LEX, we denote the version based on the XPRESS by A-LEX^X and the version which uses the algorithm ZEBRA by A-LEX^Z.

The computational study was carried out on an Intel (R) Core TM i7-3610 QM CPU with four 2.3 GHz cores each composed of two threads (although C++ code used just one thread) and 8 GB RAM.

3.1. Benchmarks. Three sets of testing problems organized by the size were used to perform the computational study. In all cases, customers' sites are considered to be also possible facility locations; that is, the sets I and J are identical. As there are no standard test problems for the facility location problem with the lexicographic minimax objective, we used the problems originally proposed for the capacitated p-median problem while interpreting the demands as b_j values (multiplicities of customers). Twenty small instances *pmedcap1*–*pmedcap20* were taken from the OR-library [36]. We also included into this set the smallest testing problem *SJC*1 used in [37]. Three larger problems taken from the same source, *SJC*2, *SJC*3, and *SJC*4, together with two instances derived from the network of 737 Spanish cities [38] constitute the medium-sized instances. The largest test problems include the problem *p*3038 originally proposed for the TSP [39] and later adjusted to the capacitated p-median problem [37]. Furthermore, considering the population data as b_j values, we created large-sized benchmarks from the road network of the Slovak Republic [40] and the road network of six southeastern US states [41] (for more details see the caption in Figure 2). All the benchmarks and the source code are available for download as a supplemental material (http://frdsa.uniza.sk/~buzna/).

3.2. Numerical Experiments. We summarized the computational results in Tables 1, 2, and 3. The abbreviations in Tables 1–3 have the following meanings:

instance: the problem name,

$|I|$: number of facility/customer locations,

p: number of facilities to be located,

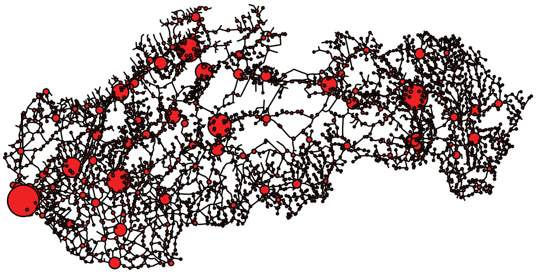
$k_{\max}$: number of D_k values,

Table 1: Computational results for the algorithm A-LEX: small instances.

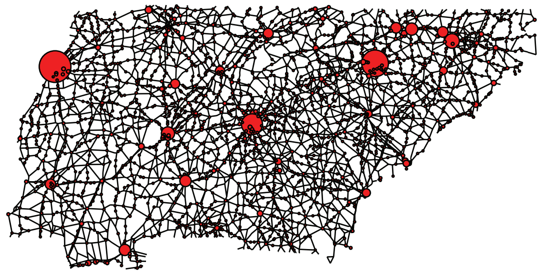
Instance	$\|I\|$	p	$k_{\max}$	O-LEX Time [s]	O-LEX k_s	A-LEX^X Time [s]	A-LEX^X k_s	Δ
*pmedcap*1	50	5	30	1.88	11	0.64	11	0
*pmedcap*2	50	5	42	3.23	8	0.67	7	0
*pmedcap*3	50	5	28	1.60	9	0.46	7	0
*pmedcap*4	50	5	43	2.97	15	0.91	15	0
*pmedcap*5	50	5	33	2.11	13	0.74	10	0
*pmedcap*6	50	5	38	2.51	11	1.21	12	0
*pmedcap*7	50	5	40	3.15	13	0.98	13	0
*pmedcap*8	50	5	39	2.75	7	0.80	6	0
*pmedcap*9	50	5	36	2.80	9	0.80	8	0
*pmedcap*10	100	10	30	1.98	10	0.59	10	0
*pmedcap*11	100	10	21	4.29	2	1.04	1	0
*pmedcap*12	100	10	26	5.21	4	1.74	3	0
*pmedcap*13	100	10	30	5.65	2	1.76	1	0
*pmedcap*14	100	10	24	4.89	3	1.91	2	0
*pmedcap*15	100	10	27	6.21	2	1.89	1	0
*pmedcap*16	100	10	23	4.59	1	1.20	0	0
*pmedcap*17	100	10	25	6.83	2	1.50	1	2
*pmedcap*18	100	10	25	4.67	2	1.70	1	0
*pmedcap*19	100	10	30	6.83	2	2.46	1	0
*pmedcap*20	100	10	20	5.95	3	0.90	2	0
*SJC*1	100	5	732	47.90	588	28.54	569	0
*SJC*1	100	10	426	22.99	283	10.82	282	0
*SJC*1	100	15	427	24.51	261	10.65	269	0
*SJC*1	100	20	150	5.31	64	2.91	65	0

Table 2: Computational results for the algorithm A-LEX: medium instances.

Instance	$\|I\|$	p	$k_{\max}$	O-LEX Time [s]	O-LEX k_s	A-LEX^X Time [s]	A-LEX^X k_s	Δ
*SJC*2	200	10	426	131.4	238	50.9	237	0
*SJC*2	200	20	306	64.4	132	37.4	128	0
*SJC*2	200	30	218	32.2	79	17.4	71	0
*SJC*2	200	40	169	20.3	39	9.7	40	0
*SJC*3	300	15	445	461.6	207	357.7	189	0
*SJC*3	300	30	267	145.1	70	68.8	58	0
*SJC*3	300	45	226	71.1	46	37.9	41	0
*SJC*3	300	60	215	53.3	50	29.8	42	0
*SJC*4	402	20	461	1371.2	161	1205.8	140	0
*SJC*4	402	40	342	1207.5	74	1052.5	58	0
*SJC*4	402	60	229	158.7	33	87.2	29	0
*SJC*4	402	80	193	144.9	25	56.2	24	0
Spain_737_1	737	37	467	116838.0	92	81185.1	65	0
Spain_737_1	737	50	348	196000.0	53	27296.2	49	0
Spain_737_1	737	185	108	12367.4	5	279.5	5	0
Spain_737_1	737	259	59	430.4	2	32.2	2	0
Spain_737_2	737	37	467	35590.7	88	29185.6	65	0
Spain_737_2	737	50	348	64005.7	59	27806.1	38	0
Spain_737_2	737	185	108	3182.3	5	232.4	5	0
Spain_737_2	737	259	59	72.5	1	43.2	1	2



(a)



(b)

FIGURE 2: Road networks used to test the algorithm A-LEX. All inhabited nodes (marked red) are considered as aggregate customers and possible locations of facilities. The diameter of nodes is scaled proportionally to the population. (a) The road network of the Slovak Republic consists of $|J| = 2928$ aggregate customers representing the locations of $|\bar{J}| = 5\,384\,909$ individuals. (b) The joined road network of six southeastern US states (Tennessee, North Carolina, South Carolina, Georgia, Alabama, and Mississippi). All inhabitants included in the dataset are considered to be customers ($|J| = 2398$, $|\bar{J}| = 14\,830\,101$).

TABLE 3: Computational results for the algorithm A-LEX: large instances.

Instance	$\|I\|$	p	k_{max}	A-LEX^Z Time [s]	A-LEX^Z k_s
$p3038$	3038	2500	33	2520.8	1
$p3038$	3038	2000	35	4204.3	2
$p3038$	3038	1500	38	8915.9	2
$p3038$	3038	900	73	190092.9	7
$p3038$	3038	700	110	17902.9	8
$p3038$	3038	100	*	*	*
$p3038$	3038	50	*	*	*
$p3038$	3038	10	1188	201165.3	36
SR	2928	2500	3	786.8	0
SR	2928	2000	4	1021.6	0
SR	2928	1500	5	1083.4	0
SR	2928	1000	8	1622.9	0
SR	2928	900	9	1988.5	0
SR	2928	700	11	2954.2	0
SR	2928	100	48	9624.7	0
SR	2928	50	65	10509.8	0
SR	2928	10	92	11888.4	0
US	2398	2000	10	1006.6	0
US	2398	1500	15	1203.7	0
US	2398	1000	22	1702.7	0
US	2398	900	25	1694.7	0
US	2398	700	33	11022.5	0
US	2398	100	*	*	*
US	2398	50	*	*	*
US	2398	10	219	9038.2	0

Table cells filled with the symbol "*" indicate the instances when the algorithm did not terminate within 3 days.

time: CPU time (in seconds) used to solve the problem,

k_s: number of skipped k values for which we did not solve the problem P_k because $g^k(\mathbf{x}^{\mathbf{k-1}}) = 0$,

Δ: number of facilities that the algorithms placed differently, calculated using the formula:

$$\Delta = \sum_{i \in I} \left| y_i^{\text{O-LEX}} - y_i^{\text{A-LEX}^X} \right|. \quad (15)$$

Due to problems with the computer memory, XPRESS solver was not able to solve large instances successfully. Therefore, in Table 3 we show the results obtained by the solver ZEBRA only. Comparison of results reveals that the algorithm A-LEX^X outperforms the algorithm O-LEX on all tested instances. A-LEX^X computed all small instances in time which accounts for 42.5% and all medium instances for 47.6% of the time needed by the algorithm O-LEX. In order to compare the quality of the solution, we evaluate

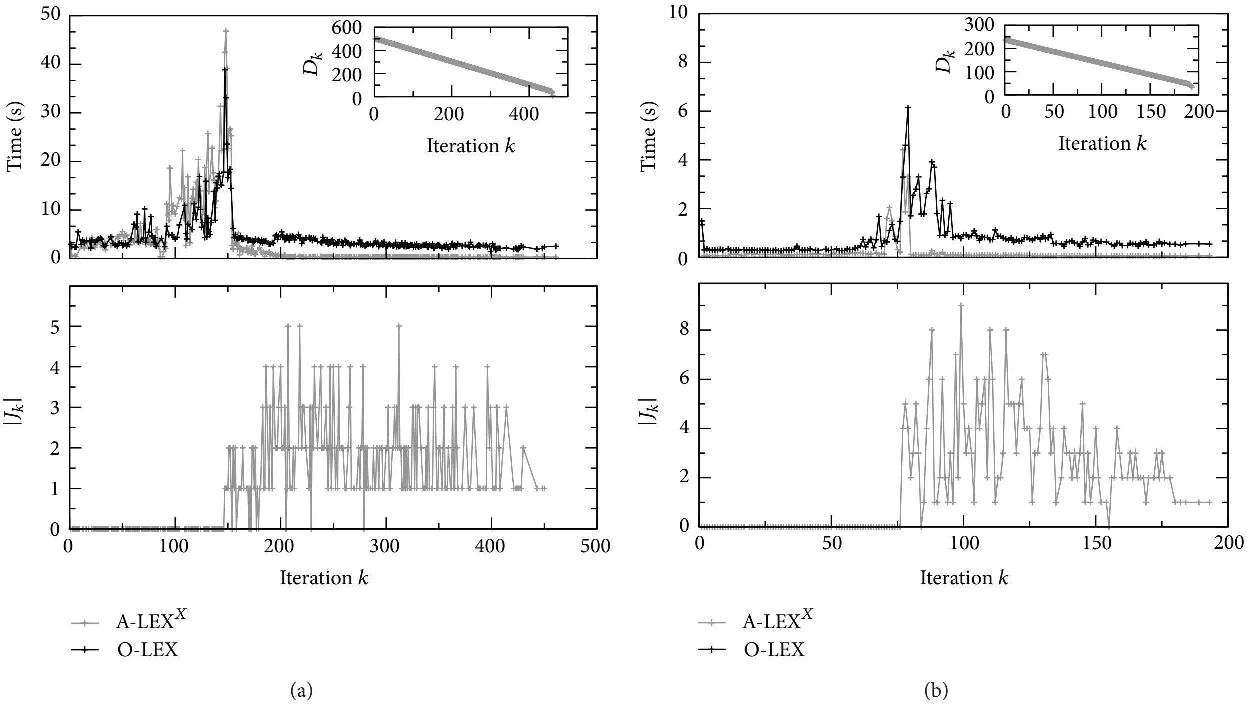


Figure 3: Computational time in seconds and the number of fixed aggregate customers $|J_k|$ as a function of the iteration step k. The results were obtained for the benchmark *SJC*4 in (a) $p = 20$ and in (b) $p = 80$. The values $|J_k|$ are identical for both algorithms, and the datasets are overlapping.

the Manhattan distance (15) between the location vectors $\mathbf{y}$. Occasionally, we find small differences in the location of facilities. However, we checked and we found that customers have the same distance to the closest located facility in all tested instances. Thus, the algorithm A-LEX found an optimal solution in all cases where we were able to compare it with the algorithm O-LEX.

To explain why the algorithm A-LEX^X is faster than the algorithm O-LEX, we need to take more detailed view on both algorithms. In the algorithm A-LEX^X, the structure of constraints in P_k is independent on k and only the structure of coefficients r_{ij}^k varies with the step k. These coefficients may take three possible values for a given customer (see the expression (13)). On the contrary, the algorithm O-LEX has simpler structure of the objective function but a new constraint is added at each iteration. New constraints make sure that the objective function values reached in the previous iterations cannot deteriorate in the following iterations [33]. Thus, the algorithm A-LEX^X outperforms O-LEX if IP solver can handle more easily the constant set of constraints with a more complex objective function and, vice versa, the algorithm O-LEX is faster if the IP solver can process faster simpler structure of the objective function with the growing set of constraints.

In Figure 3, we show the computational time needed to solve the individual instances of the problem P_k. For small values of k, where no customers are fixed (i.e., $J_k = \emptyset$), both algorithms perform comparably well. For larger values of k, the algorithm A-LEX^X systematically outperforms the algorithm O-LEX. To gain more insights into the solving process, we plotted in Figure 4 the number of branch-and-bound nodes processed by the IP solver and the number of simplex iterations required to solve the optimization problem in the root node of the branch-and-bound method. We found that the algorithm A-LEX^X needs a much smaller number of simplex iterations than the algorithm O-LEX^X. We conjecture that the number of simplex iterations is smaller in the algorithm A-LEX^X because the large values of coefficients r_{ij}^k for $d_{ij} > D_k$ allow excluding the variables taking zero values in the optimal solution faster from the basic feasible solution.

In Figure 3, we can also see that few iterations preceding and few iterations following the value k, where the first customer is fixed, take the largest portion of the computational time. This increase can be explained by the need to search through a larger number of nodes in the searching tree before the optimal integer solution is found (see Figure 4). From this point of view it could be beneficial to use a searching scheme in the algorithm A-LEX to find the value of k when the first customer is fixed and then to continue by incrementing k sequentially. This could reduce the number of iterations and could avoid processing some time-demanding iterations. In Figures 3 and 4, we showed the results obtained for the selected benchmark *SJC*4. However, it should be noted that we found qualitatively similar results with all other benchmarks.

As the algorithm A-LEX preserves the problem P_k in the form of the p-median problem, we can replace the general

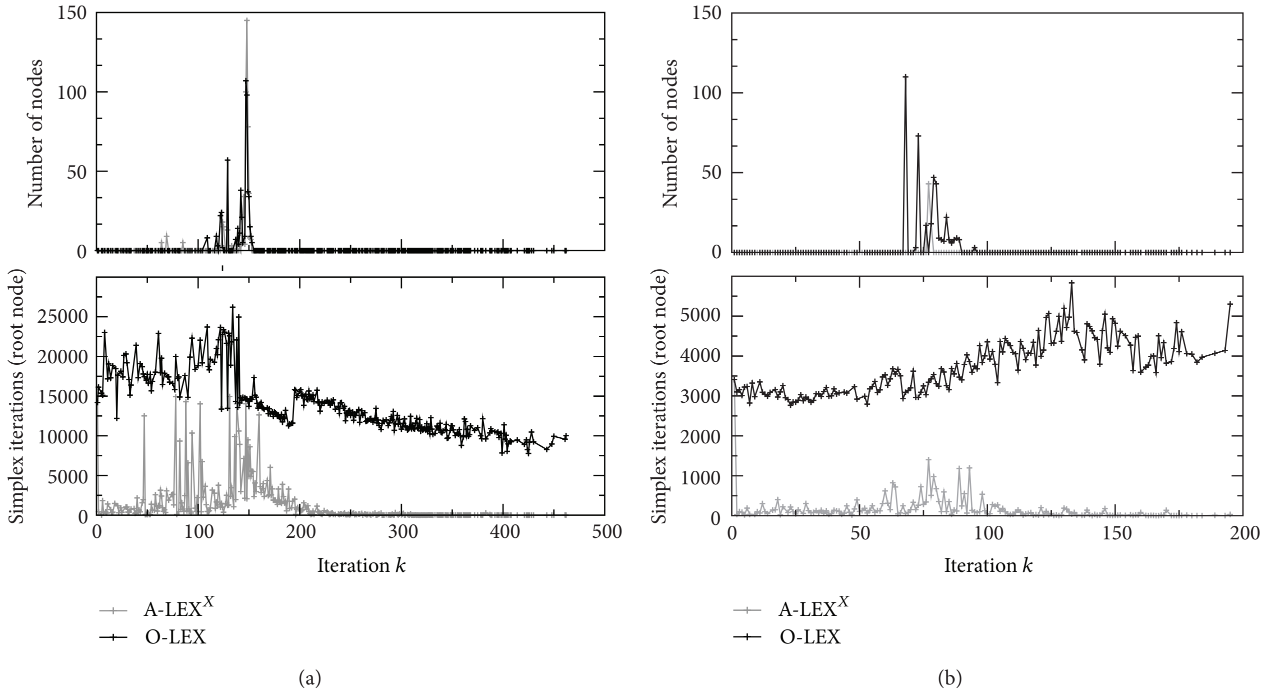


Figure 4: The number of nodes processed by the branch-and-bound method and the number of simplex iterations needed to process the root node as a function of the iteration step k. The results were obtained for the benchmark $SJC4$ in (a) $p = 20$ and in (b) $p = 80$.

purpose IP solver by the customized p-median solver. To solve the p-median problem, we used the algorithm ZEBRA [44]. The size of instances in Table 3 approaches the size of large p-median instances used to test the performance of the algorithm ZEBRA [44]. Due to the problems with the computer memory, the algorithms O-LEX and A-LEX^X could not be used to solve this set of problems. Facility location problems are strategic planning problems and the computational time is not necessarily a core issue. Therefore we set the time limit for all computations to 3 days. The problems, where p is large, can be solved relatively fast. The computational time rapidly grows by lowering the value p. We were not able to solve the problems $p3038$ and US for the intermediate p values (50 and 100) within the time limit of 3 days. Please note that similar limitations of the algorithm ZEBRA are also reported in [44].

4. Conclusions

The algorithm A-LEX, proposed in this paper, preserves the structure of constraints in the form of the p-median problem, which allows solving larger instances of problems than can be solved by the algorithm O-LEX. Moreover, the computational experiments showed that the algorithm A-LEX provides high-quality solutions. Therefore, it can be concluded that A-LEX is competitive with the existing state of the art algorithm O-LEX for solving the facility location problem with the lexicographic minimax objective function.

The proposed approximation approach is also applicable to other types of similar combinatorial optimization problems with lexicographic minimax objective. Values assigned to individual customers at different stages of the algorithm need to be included in the objective function so that we can construct rules which allow detecting whether customers can be assigned equal, smaller, or larger value than the outcome value tested on a given stage of the algorithm. To gain a computational advantage compared to all purpose IP solvers, it is needed that, for a given problem, we can use a customized exact algorithm to find the system optimum. Concrete example where this approach could be used is the maximum generalized assignment problem.

We compared the approximation algorithm A-LEX to the exact algorithm O-LEX. We are aware of the fact that it is not a standard practice to compare exact algorithms with heuristics in terms of the computational time. However, we believe that this decision can be well justified by the similarity of both algorithms, and such comparison highlights better the advantages of our approach. To our best knowledge there are no attempts in the literature to construct (meta) heuristics for the facility location problem with the lexicographic minimax objective, and such comparison could be considered as a topic for future research.

Moreover, the algorithm A-LEX can be directly applied to an arbitrary utility function dependent on the distance; therefore, it allows comparing different measures expressing how customers perceive the suitability of facility locations. The more challenging task consists of finding out how the algorithm could be extended to the composite measures representing the customer's choice behavioural model, for example, how to compute the lexicographic optimum when customers are not assigned to the closest facility, but their demand is distributed within a subset of k-nearest located facilities.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to Sergio García for making the code of the algorithm ZEBRA publicly available and for his valuable comments which helped them run the code. They thank P. Cenek, Ľ. Jánošíková, and M. Kvet for thorough reading of the paper and useful suggestions. This work was supported by the Research Grants VEGA 1/0339/13 "Advanced Microscopic Modelling and Complex Data Sources for Designing Spatially Large Public Service Systems" and APVV-0760-11 "Designing Fair Service Systems on Transportation Networks."

References

[1] H. A. Eiselt and V. Marianov, *Foundations of Location Analysis, International Series in Operations Research and Management Science*, Springer Science + Business, Berlin, Germany, 2011.

[2] M. S. Daskin, *Network and Discrete Location: Models, Algoritmhs and Applications*, John Wiley & Sons, New York, NY, USA, 1995.

[3] Z. Drezner, *Facility Location: A Survey of Applications and Methods*, Springer, New York, NY, USA, 1995.

[4] S. Müller, K. Haase, and S. Kless, "A multiperiod school location planning approach with free school choice," *Environment and Planning A*, vol. 41, no. 12, pp. 2929–2945, 2009.

[5] H. Luss, *Equitable Resources Allocation: Models, Algorithms and Applications*, John Wiley & Sons, New York, NY, USA, 2012.

[6] H. Luss, "On equitable resource allocation problems: a lexicographic minimax approach," *Operations Research*, vol. 47, no. 3, pp. 361–378, 1999.

[7] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, New York, NY, USA, 2 edition, 1992.

[8] R. Srikant, *The Mathematics of Internet Congestion Control*, Birkhäuser, Boston, Mass, USA, 2004.

[9] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, San Francisco, Calif, USA, 2004.

[10] L. R. Ford Jr. and D. R. Fulkerson, "A simple algorithm for finding maximal network flows and an application to the Hitchcock problem," *Canadian Journal of Mathematics. Journal Canadien de Mathématiques*, vol. 9, pp. 210–218, 1957.

[11] J. Rawls, *A Theory of Justice*, Harvard University Press, Cambridge, Mass, USA, 1971.

[12] N. Megiddo, "Optimal flows in networks with multiple sources and sinks," *Mathematical Programming*, vol. 7, pp. 97–107, 1974.

[13] N. Megiddo, "A good algorithm for lexicographically optimal flows in multi-terminal networks," *Bulletin of the American Mathematical Society*, vol. 83, no. 3, pp. 407–409, 1977.

[14] D. Nace and M. Pióro, "Max-min fairness and its applications to routing and load-balancing in communication networks: a tutorial," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.

[15] R. Carvalho, L. Buzna, W. Just, D. Helbing, and D. K. Arrowsmith, "Fair sharing of resources in a supply network with constraints," *Physical Review E-Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 4, Article ID 046101, 2012.

[16] D. Nace, L. N. Doan, O. Klopfenstein, and A. Bashllari, "Max-min fairness in multi-commodity flows," *Computers & Operations Research*, vol. 35, no. 2, pp. 557–573, 2008.

[17] W. Ogryczak and T. Śliwiński, "On direct methods for Lexicographic Min-Max optimization," in *Proceedings of the Computational Science and Its Applications (ICCSA '06)*, M. Gavrilova, O. Gervasi, V. Kumar et al., Eds., vol. 3982 of *Lecture Notes in Computer Science*, pp. 802–811, Springer, Berlin, Germany, 2006.

[18] W. Ogryczak and T. Śliwiński, "Sequential algorithms for exact and approximate max-min fair bandwidth allocation," in *Proceedings of the Telecommunications Network Strategy and Planning Symposium (NETWORKS '12)*, pp. 1–6, 2012.

[19] M. Zukerman, M. Mammadov, L. Tan, I. Ouveysi, and L. L. H. Andrew, "To be fair or efficient or a bit of both," *Computers and Operations Research*, vol. 35, no. 12, pp. 3787–3806, 2008.

[20] J. N. Hooker and H. P. Williams, "Combining equity and utilitarianism in a mathematical programming model," *Management Science*, vol. 58, no. 9, pp. 1682–1693, 2012.

[21] F. P. Kelly, A. K. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[22] D. Bertsimas, V. F. Farias, and N. Trichakis, "The price of fairness," *Operations Research*, vol. 59, no. 1, pp. 17–31, 2011.

[23] D. Bertsimas, V. F. Farias, and N. Trichakis, "On the efficiency-fairness trade-off," *Management Science*, vol. 58, no. 12, pp. 1–17, 2012.

[24] S. L. Hakimi, "Optimum distribution of switching centers in a communication network and some related graph theoretic problems," *Operations Research*, vol. 13, pp. 462–475, 1965.

[25] M. T. Marsh and D. A. Schilling, "Equity measurement in facility location analysis: a review and framework," *European Journal of Operational Research*, vol. 74, no. 1, pp. 1–17, 1994.

[26] T. Drezner, Z. Drezner, and J. Guyse, "Equitable service by a facility: minimizing the Gini coefficient," *Computers and Operations Research*, vol. 36, no. 12, pp. 3240–3246, 2009.

[27] W. Ogryczak, "Inequality measures and equitable locations," *Annals of Operations Research*, vol. 167, pp. 61–86, 2009.

[28] R. Z. Farahani, M. SteadieSeifi, and N. Asgari, "Multiple criteria facility location problems: a survey," *Applied Mathematical Modelling. Simulation and Computation for Engineering and Environmental Systems*, vol. 34, no. 7, pp. 1689–1709, 2010.

[29] M. Ehrgott, H. W. Hamacher, and S. Nickel, "Geometric methods to solve max-ordering location problems," *Discrete Applied Mathematics. The Journal of Combinatorial Algorithms, Informatics and Computational Sciences*, vol. 93, no. 1, pp. 3–20, 1999.

[30] A. Suzuki and Z. Drezner, "The minimum equitable radius location problem with continuous demand," *European Journal of Operational Research*, vol. 195, no. 1, pp. 17–30, 2009.

[31] O. Berman, Z. Drezner, A. Tamir, and G. O. Wesolowsky, "Optimal location with equitable loads," *Annals of Operations Research*, vol. 167, pp. 307–325, 2009.

[32] J. Krarup and P. M. Pruzan, "Reducibility of minimax to minisum 0 − 1 programming problems," *European Journal of Operational Research*, vol. 6, no. 2, pp. 125–132, 1981.

[33] W. Ogryczak, "On the lexicographic minimax approach to location problems," *European Journal of Operational Research*, vol. 100, no. 3, pp. 566–585, 1997.

[34] W. Ogryczak, "On the distribution approach to location problems," *Computers and Industrial Engineering*, vol. 37, no. 3, pp. 595–612, 1999.

[35] http://www.fico.com/.

[36] J. E. Beasley, "OR-Library: distributing test problems by electronic mail," *Journal of the Operational Research Society*, vol. 41, no. 11, pp. 1069–1072, 1990.

[37] L. A. Lorena and E. L. Senne, "A column generation approach to capacitated p-median problems," *Computers & Operations Research*, vol. 31, no. 6, pp. 863–876, 2004.

[38] J. A. Díaz and E. Fernández, "Hybrid scatter search and path relinking for the capacitated p-median problem," *European Journal of Operational Research*, vol. 169, no. 2, pp. 570–585, 2006.

[39] G. Reinelt, *The Traveling Salesman: Computational Solutions for TSP Applications*, vol. 840, Springer, Berlin, Germany, 1994.

[40] Data describing the road network of the Slovak Republic were purchased from the publisher MAPA Slovakia Plus s.r.o., 2014, http://www.mapaslovakia.sk/.

[41] United States Census Database 2000, 2014, http://www.nationalatlas.gov/mld/ce2000t.html.

[42] Y. Zhang, O. Berman, and V. Verter, "The impact of client choice on preventive healthcare facility network design," *OR Spectrum. Quantitative Approaches in Management*, vol. 34, no. 2, pp. 349–370, 2012.

[43] T. Drezner and Z. Drezner, "The gravity p-median model," *European Journal of Operational Research*, vol. 179, no. 3, pp. 1239–1251, 2007.

[44] S. García, M. Labbé, and A. Marín, "Solving large p-median problems with a radius formulation," *INFORMS Journal on Computing*, vol. 23, no. 4, pp. 546–556, 2011.

[45] 2014, https://sites.google.com/site/sergiogarciaquiles/Home.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 562079, 12 pages
http://dx.doi.org/10.1155/2014/562079

Research Article

Fair Optimization of Video Streaming Quality of Experience in LTE Networks Using Distributed Antenna Systems and Radio Resource Management

Elias Yaacoub[1] and Zaher Dawy[2]

[1] *Qatar Mobility Innovations Center (QMIC), Qatar Science and Technology Park, P.O. Box 210531, Doha, Qatar*
[2] *Electrical and Computer Engineering Department, Faculty of Engineering and Architecture, American University of Beirut, Riad El-Solh, P.O. Box 11-0236, Beirut 1107 2020, Lebanon*

Correspondence should be addressed to Elias Yaacoub; eliasy@qmic.com

Received 30 January 2014; Accepted 7 May 2014; Published 25 May 2014

Academic Editor: Dritan Nace

Network quality of experience (QoE) metrics are proposed in order to capture the overall performance of radio resource management (RRM) algorithms in terms of video quality perceived by the end users. Metrics corresponding to average, geometric mean, and minimum QoE in the network are measured when Max C/I, proportional fair, and Max-Min RRM algorithms are implemented in the network. The objective is to ensure a fair QoE for all users in the network. In our study, we investigate both the uplink (UL) and downlink (DL) directions, and we consider the use of distributed antenna systems (DASs) to enhance the performance. The performance of the various RRM methods in terms of the proposed network QoE metrics is studied in scenarios with and without DAS deployments. Results show that a combination of DAS and fair RRM algorithms can lead to significant and fair QoE enhancements for all the users in the network.

1. Introduction

With the increased video traffic in state-of-the art cellular networks, it is imperative to enhance the quality of service (QoS) of video transmissions, usually represented by the video peak signal to noise ratio (PSNR). On the other hand, video quality of experience (QoE) is gaining significant interest as a method to quantify the multimedia experience of mobile users; for example, see [1]. It can be considered as a "perceived QoS," and reflects better than QoS the quality of the video as seen by the mobile users. QoE measures are based on subjective assessment of video quality by the users. Mean opinion scores (MOS) are then collected and analyzed in order to derive an objective QoE metric translated into a mathematical formula similarly to QoS.

Most of the QoE investigations in the literature, for example, [1–4], consider link level QoE, that is, the QoE perceived by a given user in the network. The novelty in this work is in proposing metrics for assessing the QoE performance over the whole network, taking into account fairness constraints in the QoE perceived by different users. Furthermore, we study the impact of different radio resource management (RRM) algorithms on optimizing the network QoE performance and ensuring fairness towards the various users in the network.

The investigation is performed under the framework of the long term evolution (LTE) system. In LTE, orthogonal frequency division multiple access (OFDMA) is the access scheme for the downlink (DL), that is, the direction of transmission from the BS to the users. In the LTE uplink (UL), that is, the direction of transmission from the users to the BS, single carrier frequency division multiple access (SCFDMA), a modified form of OFDMA, is used [5]. In LTE, the available spectrum is divided into resource blocks (RBs), each consisting of 12 adjacent subcarriers. The assignment of an RB takes place every 1 ms, agreed to be the duration of one transmission time interval (TTI), or the duration of two

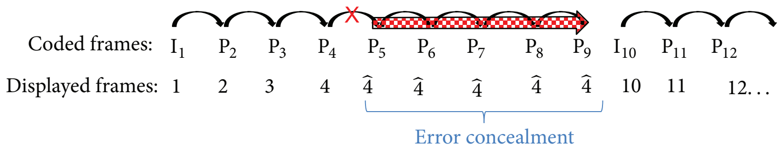


Figure 1: Error concealment example.

0.5 ms slots [6]. The LTE standard imposes the UL constraint that the RBs allocated to a single user should be consecutive with equal power allocation over the RBs [5–7].

An important method to boost performance in cellular systems is the deployment of distributed antenna systems (DASs). DASs are used to increase the coverage and capacity of wireless networks in a cost effective way. Although they are used in previous generations of wireless systems, DASs are receiving significant research attention for their deployment in LTE; for example, see [8]. The impact of DAS on LTE uplink scheduling was studied in [9], and their uplink and downlink performance were studied in [10] in the context of public safety networks.

Generally, a DAS system consists of a single central BS connected to several remote antenna heads (RAHs) distributed throughout the cell area. The BS controlling the RAHs could be colocated with any of the RAHs or in a separate location. The BS could be connected to RAHs via wired cables (e.g., fiber optic). Connection topologies include star, chain, tree, and ring topologies [11]. Since an RAH is composed of a remote antenna connected to the BS, this allows centralized control to be performed by the BS as in the conventional case while the RAHs allow extended coverage and/or more user capacity. In addition, for fixed coverage and user capacity, the RAHs provide the users with better QoS since the distance from a user to the nearest RAH will be smaller than the distance to the central BS antenna in the conventional case, which leads to a higher signal to noise ratio (SNR).

Another novelty of this work is the investigation of the role of LTE DAS systems in enhancing network QoE performance, both in the UL and DL directions. LTE RRM algorithms to ensure fair QoE optimization are investigated and compared in scenarios with and without DAS deployments.

The paper is organized as follows. Video transmission and QoS/QoE metrics are overviewed in Section 2. The proposed network QoE metrics are derived in Section 3. LTE resource allocation is described in Section 4. Section 5 analyzes radio resource management in LTE with different utility functions. The simulation results are presented and analyzed in Section 6. Finally, conclusions are drawn in Section 7.

2. Video Transmission over Wireless Channels

We consider that video sequences are encoded into groups of pictures (GOPs) according to the H.264 standard using the joint scalable video model (JSVM) software. Each GOP of N_G frames is considered to consist of one I-frame and $N_G - 1$ P-frames. When a GOP is available for real-time transmission, it should be transmitted within a duration of T_{GOP}. When T_{GOP} has elapsed, all the frames that are not received are assumed lost, and the transmission of a new GOP begins. Due to the interdependencies of the video frames, the loss of a frame in a GOP leads to the loss of all subsequent frames in the GOP, until the next I-frame is received. The loss of an I-frame leads to the loss of all the frames in the GOP [12].

2.1. Video QoS Evaluation. To measure QoS in video transmission, one of the most widely used metrics is the mean-squared error distortion. Two types of distortion affect a video sequence: source distortion and loss distortion. Source distortion depends on the compression method at the source and is beyond the scope of this paper. Loss distortion corresponds to the distortion caused by lost frames during transmission over the wireless channels. Hence, in this paper, loss distortion is considered. The distortion for replacing a frame f, of dimensions $N_1 \times N_2$ pixels, with an estimated frame $\hat{f}$ can be computed as follows [12]:

$$D\left(f, \hat{f}\right) = \frac{1}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left(f\left(n_1, n_2\right) - \hat{f}\left(n_1, n_2\right)\right)^2, \quad (1)$$

where $f(n_1, n_2)$ indicates the pixel value of frame f at position (n_1, n_2). The peak signal to noise ratio (PSNR) in this case can be expressed as

$$P_{\mathrm{SNR}}\left(f, \hat{f}\right) = \frac{\left(2^B - 1\right)^2}{D\left(f, \hat{f}\right)}, \quad (2)$$

with B the number of bits used to encode a single pixel in the picture frame.

The total (or cumulative) loss distortion depends on the frame type (I-frame, P-frame) and position (in the same GOP, the loss of a P-frame f_1 leads to more distortion than the loss of a P-frame f_2 when $f_1 < f_2$). It also depends on the coding method used and error concealment method. The most common approach for error concealment is known as the previous frame concealment method. It consists of repeating the last correctly received frame until the next I-frame is received [12]. An example is shown in Figure 1 assuming frame number 5 is lost. The notation $\widehat{X}$ in Figure 1 is used to indicate that frame X is used for error concealment to replace the lost frame at a given position. In the shown example, frame number 4 would be used to replace all the remaining frames in the GOP. The total loss distortion when frame f is lost in a GOP of N_G frames with one I-frame and $N_G - 1$ P-frames, using the previous frame concealment method, can be obtained from (1) as follows:

$$D^{(f)} = \sum_{y=f}^{N_G} D(y, f-1). \tag{3}$$

The expression in (3) is obtained due to replacing all lost frames in the GOP from frame f onwards by the last correctly received frame $f-1$. When an I-frame is lost, it is replaced by the last correctly received frame from the previous GOP.

2.2. Video QoE Metric. QoE is gaining significant interest as a method to quantify the multimedia experience of mobile users; for example, see [1–4]. A survey of QoE techniques is presented in [1]. QoE tries to measure the QoS as it is finally perceived by the end user. For example, following [2], the QoE can be related to the PSNR as follows:

$$Q = \frac{1}{1 + e^{b_1(P_{\text{SNR}} - b_2)}}, \tag{4}$$

where b_1 and b_2 are parameters that depend on the video characteristics and P_{SNR} is the PSNR expressed in dB. In (4), $Q = 0$ indicates the best quality and $Q = 1$ indicates the worst quality. The relation between QoE and QoS to PSNR and distortion is an ongoing research activity; for example, see [13, 14]. In [13], subjective quality assessment of video is performed in order to determine novel QoE metrics. In [3], an empirical QoE metric taking into account PSNR, spatial resolution, and frame rate, in addition to spatial and temporal variances, was derived. A QoE metric derived in [4], based on the metric in [2], is expressed as

$$Q_m = Q_{\max}\left(1 - \frac{1}{1 + e^{b_1(P_{\text{SNR}} - b_2)}}\right) \cdot \frac{1 - e^{-b_3(f/f_{\max})}}{1 - e^{-b_3}}, \tag{5}$$

where b_3 is another parameter that depends on the video characteristics, $Q_{\max}$ is a constant corresponding to maximum quality, f is the frame rate at which the video is displayed, and $f_{\max}$ is the maximum frame rate. The QoE derivations take into account user experience while playing the video and are not inherently designed to assess the transmission over wireless channels. Using the previous frame error concealment method, the same frame rate can be maintained after error concealment (i.e., $f = f_{\max}$), and hence (5) can be simplified to

$$Q_m = Q_{\max}\left(1 - \frac{1}{1 + e^{b_1(P_{\text{SNR}} - b_2)}}\right). \tag{6}$$

Hence, in this paper, we use (6) with $Q_{\max} = 100$, thus displaying QoE on a scale from 0 to 100.

In practice, during the streaming of a stored video at the streaming server or BS, the video characteristics can be extracted offline and used to determine b_1, b_2, and b_3 in the QoE metrics. Although these parameters are difficult to extract during the streaming of a live video, an approach for their dynamic real-time estimation is presented in [4].

3. Network QoE

The QoE metric of (6) is an "individual" metric reflecting the QoE experience of a particular user. The objective of this paper is to investigate radio resource management (RRM) algorithms that ensure a fair QoE satisfaction for all users in the network. Hence, "network" QoE metrics reflecting the overall performance of RRM algorithms in terms of enhancing QoE for all users in the network need to be derived. This section presents novel network QoE metrics that could help assess the fairness of RRM algorithms in ensuring QoE satisfaction.

3.1. Proposed Network QoE Metrics. The first metric is the average QoE. It is given by

$$Q_m^{(\text{avg})} = \frac{1}{K}\sum_{k=1}^{K} Q_{m,k}, \tag{7}$$

where $Q_{m,k}$ is the QoE metric of user k, expressed as in (6) for example. The metric in (7) reflects the average performance in the network. However, in some instances, $Q_m^{(\text{avg})}$ could be relatively high when some users have very high QoE while others have a relatively low QoE. Consequently, this could mask the unfairness towards users with low QoE.

A possible solution to this problem is to derive RRM algorithms maximizing the minimum QoE in the network given by

$$Q_m^{(\min)} = \min_k Q_{m,k}. \tag{8}$$

This allows enhancing the worst case performance. However, this could come at the expense of users with good channel conditions (and who could achieve high QoE) that will be unfavored by the RRM algorithms in order to increase the QoE of worst case users.

A tradeoff between the metrics in (7) and (8) could be the use of the geometric mean QoE, given by

$$Q_m^{(\text{gm})} = \left(\prod_{k=1}^{K} Q_{m,k}\right)^{1/K}. \tag{9}$$

The metric (9) is fair, since a user with a QoE close to zero will make the whole product in $Q_m^{(\text{gm})}$ go to zero. Hence, any RRM algorithm maximizing $Q_m^{(\text{gm})}$ would avoid having any user with very low QoE. In addition, the metric (9) will reasonably favor users with good wireless channels (capable of achieving high QoE), since a high QoE will contribute in increasing the product in (9).

3.2. QoE Optimization. Using the metrics derived in (7)–(9), the objective is to maximize a network QoE metric as follows:

$$\max_{\alpha_{k_l,i,l}^{(\text{DL})}, \alpha_{k_l,i,l}^{(\text{UL})}, P_l^{(\text{DL})}, P_{k_l}^{(\text{UL})}} Q_m^{(\text{net})}, \tag{10}$$

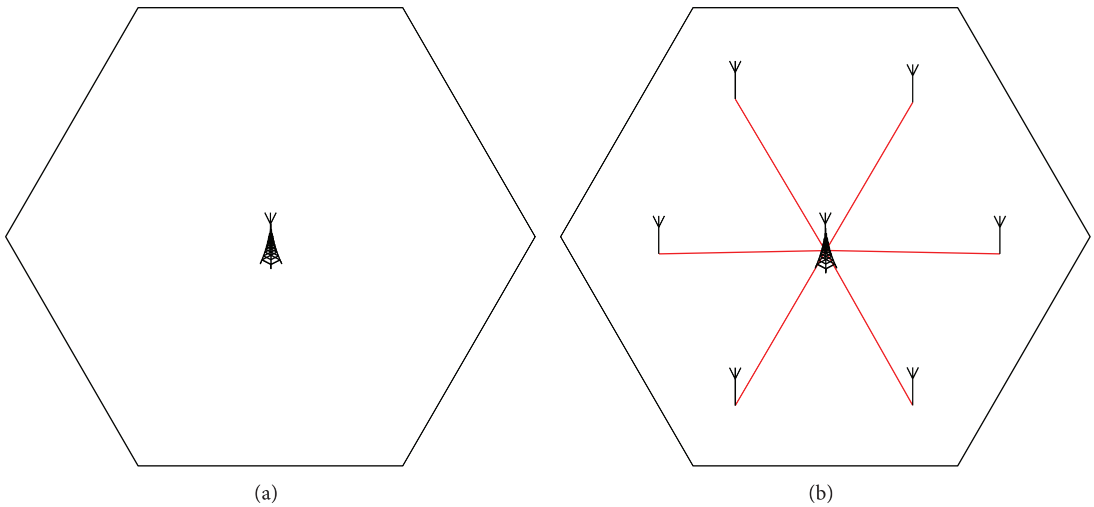


Figure 2: Deployment scenarios: (a) BS only; (b) DAS with one BS and six RAHs.

subject to

$$P_{k_l}^{(\mathrm{UL})} \leq P_{k_l,\max}^{(\mathrm{UL})}; \quad \forall k_l = 1, \ldots, K_l; \ \forall l = 1, \ldots, N_{\mathrm{BS}}, \tag{11}$$

$$P_{l}^{(\mathrm{DL})} \leq P_{l,\max}^{(\mathrm{DL})}; \quad \forall l = 1, \ldots, N_{\mathrm{BS}}, \tag{12}$$

$$\sum_{k_l=1}^{K_l} \alpha_{k_l,i,l}^{(\mathrm{UL})} \leq 1; \quad \forall i = 1, \ldots, N_{\mathrm{sub}}^{(\mathrm{UL})}; \ \forall l = 1, \ldots, N_{\mathrm{BS}}, \tag{13}$$

$$\sum_{k_l=1}^{K_l} \alpha_{k_l,i,l}^{(\mathrm{DL})} \leq 1; \quad \forall i = 1, \ldots, N_{\mathrm{sub}}^{(\mathrm{DL})}; \ \forall l = 1, \ldots, N_{\mathrm{BS}}, \tag{14}$$

where $Q_m^{(\mathrm{net})}$ is one of the metrics in (7)–(9). In addition, K_l is the number of users in cell l, N_{BS} is the number of base stations (BSs), $P_{k_l}^{(\mathrm{UL})}$ is the transmit power of user k_l in cell l in the uplink (UL), $P_{k_l,\max}^{(\mathrm{UL})}$ is its maximum transmit power, $P_l^{(\mathrm{DL})}$ is the transmit power of BS l in the downlink (DL), and $P_{l,\max}^{(\mathrm{DL})}$ is its maximum transmit power. Furthermore, $N_{\mathrm{sub}}^{(\mathrm{UL})}$ and $N_{\mathrm{sub}}^{(\mathrm{DL})}$ are the numbers of OFDMA subcarriers in the UL and DL, respectively. Finally, $\alpha_{k_l,i,l}^{(\mathrm{UL})}$ and $\alpha_{k_l,i,l}^{(\mathrm{DL})}$ are indicator variables for the UL and DL, respectively. They are set to one if subcarrier i is allocated to user k_l in cell l and set to zero otherwise.

The constraints in (11) and (12) indicate that the transmit power cannot exceed the maximum power for the UL and DL, respectively. The constraints in (13) and (14) correspond to the exclusivity of subcarrier allocations in each cell for the UL and DL, respectively, since, in each cell, a subcarrier can be allocated at most to a unique user at a given scheduling instant.

It should be noted that, using individual QoS metrics instead of individual QoE metrics in (7)–(9), these equations become novel definitions of network QoS metrics instead of network QoE metrics. Indeed, Section 6.2 presents a performance comparison when network QoS (considered to be video PSNR in this paper) and network QoE metrics are used.

4. LTE Resource Allocation

In this Section, we describe the approach used in this paper for LTE resource allocation.

4.1. System Model. The system model is displayed in Figure 2, where Scenario (a) shows a traditional BS at the cell center, whereas Scenario (b) shows a DAS deployment with six RAHs deployed throughout the cell area. Consequently, Scenario (b) consists of seven RAHs: one located at the cell center and six located at a distance of $2R_c/3$, with 60 degrees angular separation between them.

In this work, we consider a single cell, and we compare the LTE performance in the presence and absence of DASs, for the scenarios presented in Figure 2. In the comparisons, we consider the same coverage area and the same number of users in the cell in the case of a single centralized BS (Scenario (a) in Figure 2) and in the case of DASs (Scenario (b) in Figure 2). We also consider the same number of subcarriers in both scenarios.

In the DL, users stream video files coming from the BS in real time. We also consider video transmission in the UL, which could correspond, in practice, to wireless transmission of videos captured by surveillance cameras, to police, or other public safety teams sending real-time videos to a command center during an incident or pursuit, among other possible applications.

In the DAS scenario, the presence of RAHs is transparent to the users who act as if there was only a single central BS in the cell. The communication between the central BS and RAHs is via fiber optic (or microwave links having a nonoverlapping spectrum with LTE) and consequently does not consume any LTE radio resources. The presence of RAHs contributes in enhancing the channel states of the different users by providing each user with an antenna that is closer to it than the central BS antenna. Hence, the channel gain of user k over subcarrier i on the link with RAH x can be expressed as follows:

$$H_{k,i,\mathrm{dB}}^{(x)} = \left(-\kappa - \upsilon \log_{10} d_k^{(x)}\right) - \xi_{k,i} + 10\log_{10} F_{k,i}, \tag{15}$$

where the first factor captures propagation loss, with κ the path loss constant, $d_k^{(x)}$ the distance in km from mobile k to RAH x, and υ the path loss exponent. The second factor, $\xi_{k,i}$, captures log-normal shadowing with zero mean and a standard deviation σ_ξ, whereas the last factor, $F_{k,i}$, corresponds to Rayleigh fading power with a Rayleigh parameter b such that $E\{|b|^2\} = 1$.

In the traditional scenario (Scenario (a)), denoting by $x = 1$ the radio head colocated with the BS at the cell center, the channel gain between the BS and user k over subcarrier i is expressed as $H_{k,i} = H_{k,i}^{(1)}$. With DASs, it will transparently "appear" to user k that its channel gain with the BS over subcarrier i is

$$H_{k,i} = \arg\max_x H_{k,i}^{(x)}. \tag{16}$$

It should be noted that the above analysis applies to both the UL and DL, depending on whether i is an UL or DL subcarrier, respectively.

4.2. Throughput Calculations in the Uplink. Let $P_{k,i}^{(\mathrm{UL})}$ be the power transmitted by user k over subcarrier i, $P_{k,\max}^{(\mathrm{UL})}$ be the maximum transmission power of user k, and $R_k^{(\mathrm{UL})}$ be its achievable throughput in the UL. Then, the SCFDMA throughput of user k is given by

$$\begin{aligned} R_k^{(\mathrm{UL})} &\\ &= B_{\mathrm{sub}}^{(\mathrm{UL})} \left|\mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}\right| \cdot \log_2\left(1 + \beta \gamma_k^{(\mathrm{UL})}\left(\mathbf{P}_{\mathbf{k}}^{(\mathrm{UL})}, \mathcal{I}_{\mathrm{sub},k_l}^{(\mathrm{UL})}\right)\right), \end{aligned} \tag{17}$$

where $B_{\mathrm{sub}}^{(\mathrm{UL})}$ is the UL subcarrier bandwidth, $|\mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}|$ is the cardinality of $\mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}$, $N_{\mathrm{sub}}^{(\mathrm{UL})}$ is the number of UL subcarriers, and $\mathbf{P}_{\mathbf{k}}^{(\mathrm{UL})}$ represents a vector of the transmitted power on each subcarrier, $P_{k,i}$. β is called the SNR gap. It indicates the difference between the SNR needed to achieve a certain data transmission rate for a practical M-QAM system and the theoretical limit (Shannon capacity) [15]. It is given by

$$\beta = \frac{-1.5}{\ln\left(5P_b\right)}, \tag{18}$$

where P_b denotes the bit error rate (BER). Finally, $\gamma_k^{(\mathrm{UL})}(\mathbf{P}_{\mathbf{k}}^{(\mathrm{UL})}, \mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})})$ is the SNR of user k after minimum mean squared error (MMSE) frequency domain equalization at the receiver [5]:

$$\begin{aligned} &\gamma_k^{(\mathrm{UL})}\left(\mathbf{P}_{\mathbf{k}}^{(\mathrm{UL})}, \mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}\right) \\ &= \left(\frac{1}{\left(1/\left|\mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}\right|\right)\sum_{i\in\mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}}\left(\gamma_{k,i}^{(\mathrm{UL})}/\left(\gamma_{k,i}^{(\mathrm{UL})}+1\right)\right)} - 1\right)^{-1}. \end{aligned} \tag{19}$$

In (19), $\gamma_{k,i}^{(\mathrm{UL})}$ is the UL SNR of user k over subcarrier i. It is given by

$$\gamma_{k,i}^{(\mathrm{UL})} = \frac{P_{k,i}^{(\mathrm{UL})} H_{k,i}}{\sigma_{\mathrm{BS},i}^2}, \tag{20}$$

where $H_{k,i}$ is the channel gain over UL subcarrier i allocated to user k and $\sigma_{\mathrm{BS},i}^2$ is the noise power at the receiver of the BS (i.e., the receiver of the RAH that is nearest to user k).

The LTE standard imposes the constraint that the RBs allocated to a single user should be consecutive with equal power allocation over the subcarriers of those RBs [5–7]. The contiguous RB constraint is enforced by Step 4 of the algorithm in Section 4.4. To ensure equal power allocation, we set

$$P_{k,i}^{(\mathrm{UL})} = \frac{P_{k,\max}^{(\mathrm{UL})}}{\left|\mathcal{I}_{\mathrm{sub},k}^{(\mathrm{UL})}\right|}. \tag{21}$$

4.3. Throughput Calculations in the Downlink. The DL achievable throughput of user k over RB n is given by

$$R_{k,n}^{(\mathrm{DL})} = \sum_{i\in\mathrm{RB}n} B_{\mathrm{sub}}^{(\mathrm{DL})} \cdot \log_2\left(1 + \beta\gamma_{k,i}^{(\mathrm{DL})}\right), \tag{22}$$

where $B_{\mathrm{sub}}^{(\mathrm{DL})}$ is the subcarrier bandwidth. In (22), the summation is taken over the consecutive subcarriers that constitute RB n. We consider that the BS transmits at the maximum power $P_{\mathrm{BS,max}}$ and the power is assumed to be subdivided equally among all the subcarriers. Hence, the DL SNR of user k over a single subcarrier i, $\gamma_{k,i}^{(\mathrm{DL})}$, is given by

$$\gamma_{k,i}^{(\mathrm{DL})} = \frac{\left(P_{\mathrm{BS,max}}/N_{\mathrm{sub}}^{(\mathrm{DL})}\right) H_{k,i}}{\sigma_{k,i}^2}, \tag{23}$$

where $N_{\mathrm{sub}}^{(\mathrm{DL})}$ is the total number of DL subcarriers, $H_{k,i}$ is the channel gain over DL subcarrier i allocated to user k, and $\sigma_{k,i}^2$ is the noise power at the receiver of user k.

4.4. LTE Resource Allocation. The resource allocation algorithm presented in this section was proposed by the authors in [16] where it was used to enhance individual QoEs. It is repeated here for completeness of the analysis. Furthermore, it is used with various utilities in order to maximize the various novel network QoE metrics presented in Section 3. The algorithm is applicable to both UL and DL. Hence, we will drop the superscripts (UL) and (DL) to avoid repetition. We denote by $\mathcal{I}_{\mathrm{sub},k}$ the set of subcarriers allocated to user k, $\mathcal{I}_{\mathrm{RB},k}$ the set of RBs allocated to user k, N_{RB} the total number of RBs, K the number of users, and R_k the achievable throughput of user k. We define $U(R_k \mid \mathcal{I}_{\mathrm{RB},k})$ as the utility of user k as a function of the throughput R_k given the allocation $\mathcal{I}_{\mathrm{RB},k}$.

The resource allocation algorithm presented below consists of allocating RB n to user k in a way to maximize the difference

$$\Lambda_{n,k} = U\left(R_k \mid \mathcal{I}_{\mathrm{RB},k} \cup \{n\}\right) - U\left(R_k \mid \mathcal{I}_{\mathrm{RB},k}\right), \tag{24}$$

where the marginal utility, $\Lambda_{n,k}$, represents the gain in the utility function U when RB n is allocated to user k, compared to the utility of user k before the allocation of n. The algorithm is described as follows.

(i) Consider the set of available RBs $\mathcal{I}_{\text{avail\_RB}} \subseteq \{1, 2, \ldots, N_{\text{RB}}\}$ and the set of available users $\mathcal{I}_{\text{avail\_users}} \subseteq \{1, 2, \ldots, K\}$. At the start of the algorithm, $\mathcal{I}_{\text{avail\_RB}} = \{1, 2, \ldots, N_{\text{RB}}\}$ and $\mathcal{I}_{\text{avail\_users}} = \{1, 2, \ldots, K\}$.

Step 1. Find the user that has the highest marginal utility defined in (24) among all available users when the first available RB in $\mathcal{I}_{\text{avail\_RB}}$ is allocated to it. In other words, for each RB n, find the user k^* such that

$$k^* = \arg\max_k \Lambda_{n,k}. \tag{25}$$

Step 2. Allocate RB n to user k^*: $\mathcal{I}_{\text{RB},k^*} = \mathcal{I}_{\text{RB},k^*} \cup \{n\}$.

Step 3. Delete the RB from the set of available RBs:

$$\mathcal{I}_{\text{avail\_RB}} = \mathcal{I}_{\text{avail\_RB}} - \{n\}. \tag{26}$$

Step 4. This step is only for the UL direction in order to guarantee the contiguity of subcarrier allocations. It is not needed for the DL. In the UL, if k^* is the same user to which RB $n-1$ was allocated, that is, $k^* = \arg\max_k \Lambda_{n-1,k}$, then keep k^* in $\mathcal{I}_{\text{avail\_users}}$. Otherwise, delete user k^* from the set of available users:

$$\mathcal{I}_{\text{avail\_users}} = \mathcal{I}_{\text{avail\_users}} - \{k^*\}. \tag{27}$$

(ii) Repeat Steps 1, 2, 3, and 4, until there are no available RBs or no available users.

The utility function depends on the data rate and can be changed depending on the different services and QoS/QoE requirements. Different utility functions that can be used with the proposed algorithm are presented in Section 5.

5. RRM Utility Selection for QoE Maximization

To perform the maximization of (10), we use the utility maximization algorithm of Section 4.4, applicable for the UL and DL. The proposed algorithm can be applied with a wide range of utility functions, being able to achieve various objectives, with each objective represented by a certain utility function.

To explicitly implement RRM algorithms maximizing QoE metrics, real-time feedback is needed from mobile terminals about the QoE achieved by each user, which requires modifications to the standards. Furthermore, this feedback would depend on each video sequence being streamed. In this paper, we propose to perform RRM without any QoE feedback, using standard compliant algorithms. With the utility maximization algorithm of Section 4.4, we use utility functions depending on the users' data rates. We investigate Max C/I, proportional fair, and Max-Min utilities for data rates and study the impact of their implementation on the average, geometric mean, and minimum QoE network metrics.

5.1. Max C/I Utility. Letting the utility equal to the data rate $U_k = R_k$, the algorithm of Section 4.4 leads to a maximization of the sum rate of the cell (and hence of the average data rate in the cell). However, in this case, users close to the BS will be allocated most of the resources, and hence will have the highest QoE. However, edge users will generally suffer from starvation and will have very low data rates and consequently very low QoE.

5.2. Max-Min Utility. In this section, we discuss utilities corresponding to the problem of rate maximization with fairness constraints, by attempting to maximize the minimum data rate in the network, for example, [17, 18]. A vector $\mathbf{R}$ of user data rates is Max-Min fair if and only if, for each k, an increase in R_k leads to a decrease in R_j for some j with $R_j < R_k$ [17]. Max-Min utilities lead to more fairness by increasing the priority of users having lower rates [18]. It was shown that Max-Min fairness can be achieved by utilities of the form [18]:

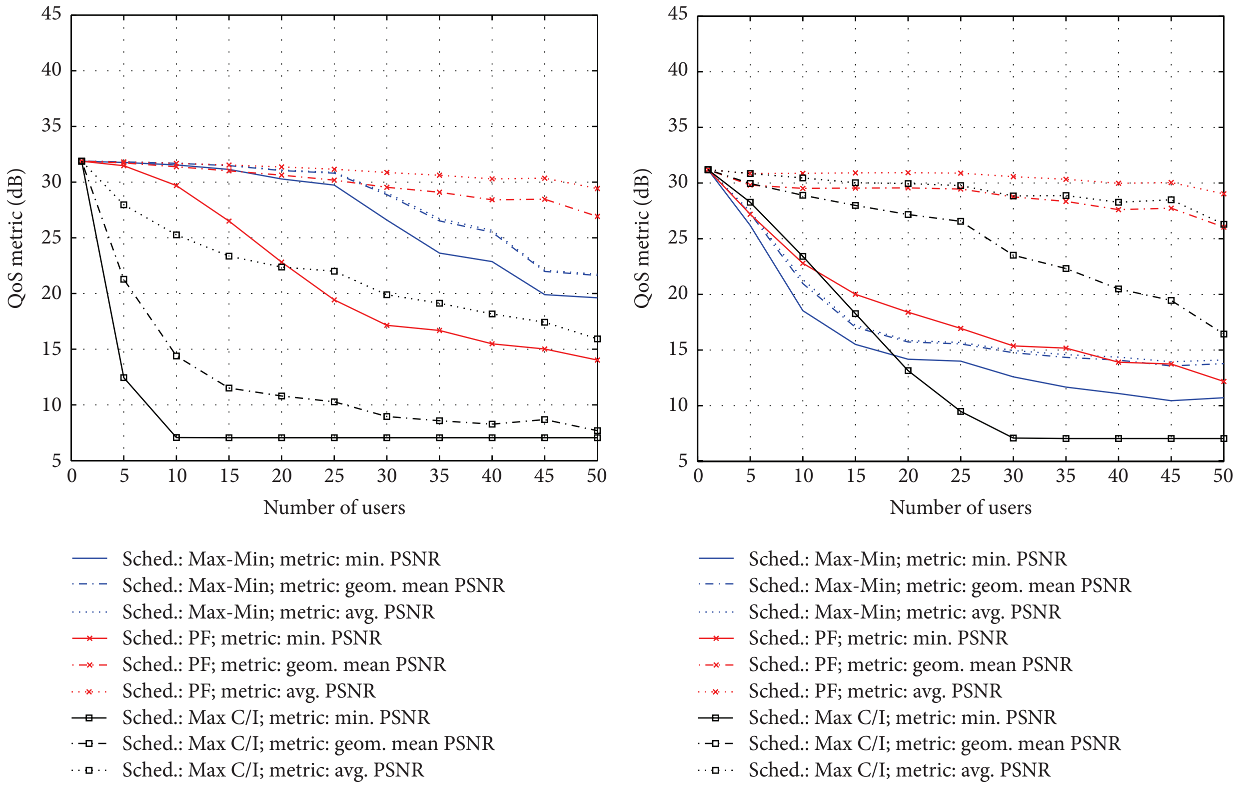
$$U_k(R_k) = -\frac{R_k^{-\alpha}}{\alpha}, \quad \alpha > 0, \tag{28}$$

where the parameter α determines the degree of fairness. Max-Min fairness is attained when $\alpha \to \infty$ [18]. We use $\alpha = 10$ in this paper.

5.3. Proportional Fair Utility. In this section, in order to ensure a more fair allocation of wireless resources, we model the problem as a bargaining game. We consider that each user is a player who wants to maximize its payoff, considered to be its data rate. Consequently, players should share the resources in an optimal way, that is, a way they cannot jointly improve on. The resources to be shared are the OFDMA subcarriers. Allocating the shared resources in a way to maximize the players' payoffs is equivalent to allocating the subcarriers to users in a way to maximize each user's data rate, given the shares of subcarriers allocated to the other users. With each user wanting to selfishly maximize its data rate, the users engage in a "bargaining" process. It is a well-known result in game theory that the solution to the bargaining problem maximizes the Nash product N_P [19]:

$$\begin{aligned} \max \prod_{k=1}^{K} R_k \Longleftrightarrow \max \ln\left(\prod_{k=1}^{K} R_k\right) \\ = \max \sum_{k=1}^{K} \ln(R_k). \end{aligned} \tag{29}$$

Interestingly, the algorithmic implementation of (29) can be handled by the algorithm of Section 4.4, by using, in that algorithm, $U_k = \ln(R_k)$ as the utility of user k, where ln represents the natural logarithm. Maximizing the sum of logarithms in (29) is equivalent to maximizing the product and is easier to implement numerically. This approach represents proportional fair (PF) scheduling, a well-known resource allocation approach in wireless communications systems. PF scheduling is known to correspond to a sum of the logarithms



(a) Downlink (b) Uplink

Figure 3: QoS metric.

of the user rates and represents the Nash bargaining solution [20]. Hence, letting $U = \ln(R)$ provides proportional fairness. Using, in the logarithm, the achievable data rate at the current scheduling instant achieves proportional fairness in frequency (PFF), whereas including the previous scheduling instants by using the cumulative data rate (since the start of the video transmission to/by the user) achieves proportional fairness in time and frequency (PFTF) [9]. In this paper, PFTF is used, since it gives a fair allocation for all users to transmit/receive their videos on the UL/DL. It can be easily shown from (29) that PF RRM algorithms maximize the geometric mean of the user data rates.

6. Results and Discussion

This section presents the simulation results obtained by comparing the scenarios of Figure 2 using RRM with the utilities of Section 5 to maximize the network QoS/QoE metrics of Section 3.

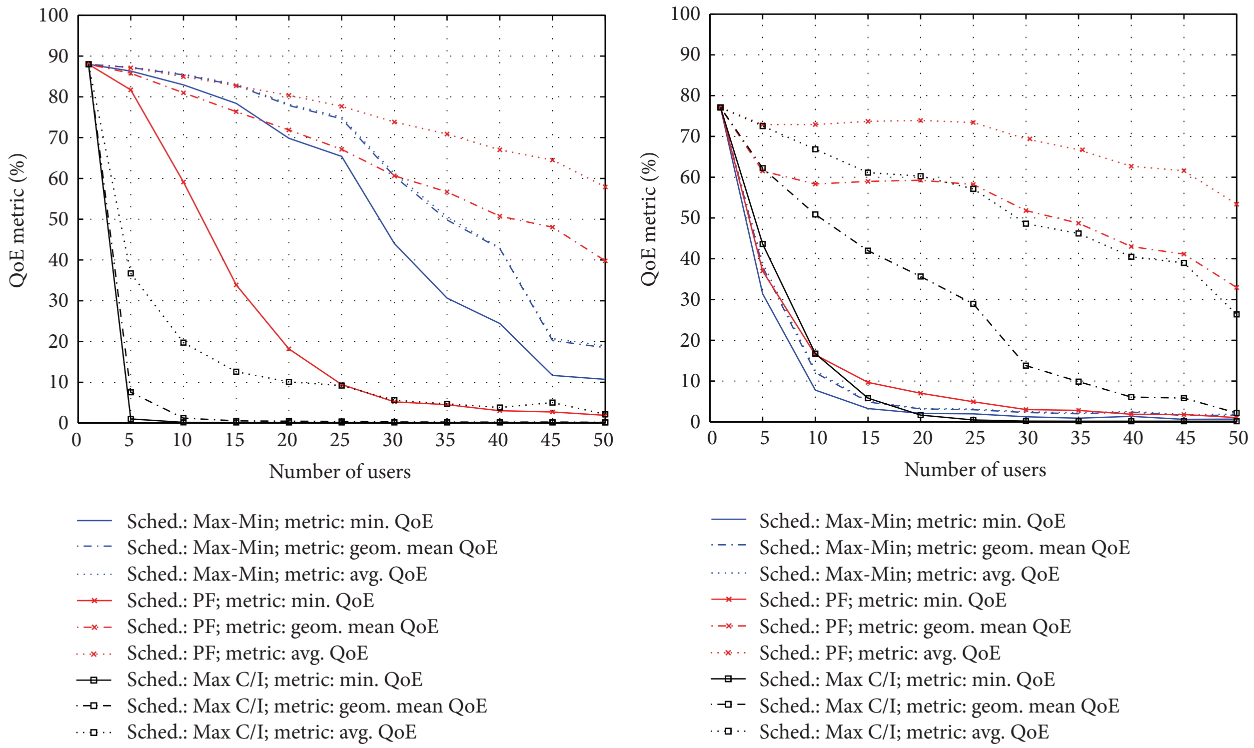
6.1. Simulation Model. The simulation model consists of a single cell with a BS equipped with an omnidirectional antenna or several RAHs each having an omnidirectional antenna, as shown in Figure 2. The simulation parameters are shown in Table 1. LTE parameters are obtained from [7, 21], and channel parameters are obtained from [22]. Users are considered to be uniformly distributed in the cell area.

Table 1: Simulation parameters.

Parameter	Value	Parameter	Value
κ	−128.1 dB	υ	3.76
σ_ξ (dB)	8 dB	Rayleigh parameter b	$E[b^2] = 1$
$B^{(\mathrm{DL})}$	5 MHz	$B^{(\mathrm{UL})}$	5 MHz
$N_{\mathrm{RB}}^{(\mathrm{DL})}$	25	$N_{\mathrm{RB}}^{(\mathrm{UL})}$	25
$B_{\mathrm{sub}}^{(\mathrm{DL})}$	15 kHz	$B_{\mathrm{sub}}^{(\mathrm{UL})}$	15 kHz
$P_{\mathrm{BS,max}}^{(\mathrm{DL})}$	5 W	$P_{k,\mathrm{max}}^{(\mathrm{UL})}$	0.125 W

To simulate the video transmission in both directions, UL and DL, the football sequence, encoded in QCIF format, is used, with GOPs consisting of 15 frames, one I-frame and 14 P-frames, having a GOP duration $T_{\mathrm{GOP}} = 1$ s. The results are averaged over 2500 iterations, where, in each iteration, a video sequence has to be transmitted from the BS to each mobile user in the DL, or from each mobile terminal to the BS in the UL.

6.2. Video QoS/QoE Results without DAS Deployment. Figures 3 and 4 show the network QoS and QoE results, respectively, for the DL and UL. Comparing the DL and UL performance in both Figures 3 and 4, it can be noted that the QoS/QoE performance in the DL is slightly better than in the UL, due to the higher transmission power available at the BS.



(a) Downlink

(b) Uplink

FIGURE 4: QoE metric.

It should be noted that although the same video sequence was used for simulation purposes, the scenario considered corresponds in practice to a unicasting scenario where different videos are transmitted to (by) each user in the DL (UL). Otherwise, it would be better to perform multicasting by the BS in the DL and collaborative transmission by the users in the UL, which represent interesting topics for future research.

It can be seen that PF scheduling maximizes the average network QoS/QoE both in the DL and UL. It also leads to more fairness in the UL, since the best results for geometric mean QoS/QoE are achieved with PF scheduling. The same is achieved in the DL when the number of users increases by maximizing the geometric mean QoS/QoE when the number of users is above 30. Max-Min scheduling is shown to maximize the minimum QoS/QoE in the downlink and to outperform proportional fair scheduling in terms of geometric mean QoS/QoE when the number of users in the DL is relatively low (below 30). However, its performance in the UL is not as good due to the limited transmit power of mobile terminals. In fact, in the UL, it is outperformed by PF for all three network QoS/QoE metrics and by Max C/I scheduling for the average and geometric mean QoS/QoE metrics. It is also outperformed by Max C/I for the average QoS/QoE metric when the number of users is below 20.

RRM using Max C/I is extremely unfair in the DL, as shown in Figures 3(a) and 4(a) for all three QoS/QoE metrics, respectively. This is due to the fact that it allocates the subcarriers and power available at the BS to users that are relatively close to the BS, which deprives users that are further away from wireless resources. Due to error propagation in video sequences caused by lost frames, this leads to very low QoS/QoE results, especially for the metrics involving a certain notion of fairness: the min QoS/QoE and geometric mean QoS/QoE. In the UL, Max C/I performs better, mainly due to the fact that the power is now distributed: each user in the UL has its own transmit power, conversely to the DL, where the power is concentrated at a single entity, the BS. However, its performance degradation is fast as the number of users increases, especially for the min QoS/QoE in the network.

Comparing Figure 3 to Figure 4, it can be seen that all combinations of {RRM algorithm, network quality metric} have the same performance trends in both figures. In other words, the same conclusions can be reached in terms of the superiority of a method over another for both QoS and QoE. However, the comparison shows that the performance degradation is faster in the QoE case. For example, the performance gap between the average network metric and the geometric mean metric appears to be reduced in the QoS case (Figure 3), whereas it looks significant in the QoE case (Figure 4). As another example, the average QoS metric is nonzero (although low) with Max C/I scheduling (Figure 3), whereas the average QoE metric goes to zero when the number of users increases with Max C/I (Figure 4). This explains the motivation behind using QoE metrics instead of QoS metrics, since they correspond to a more accurate representation of the quality perceived by the end users.

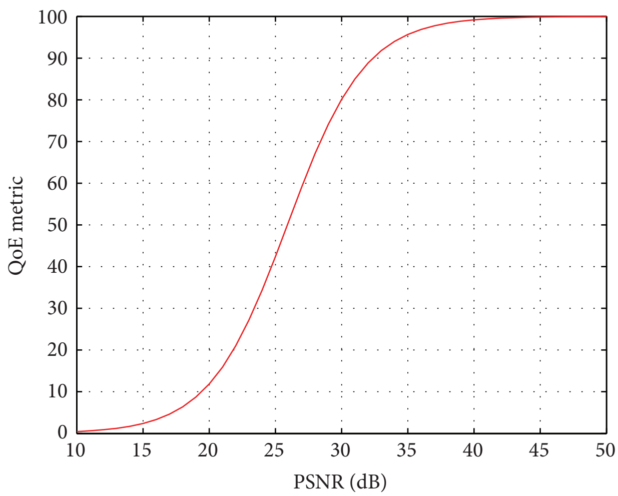


Figure 5: QoE metric versus PSNR for the football video sequence.

In fact, Figure 5 shows the plot of QoE versus QoS (PSNR) for the football video sequence used in the simulations. It can be seen that when the PSNR is too high or too low, the variation in QoE is unnoticeable. However, the sensitivity increases in the intermediate ranges. For example, a drop of 5 dB from 30 dB to 25 dB in PSNR leads to a dramatic fall from 80% to 40% in QoE.

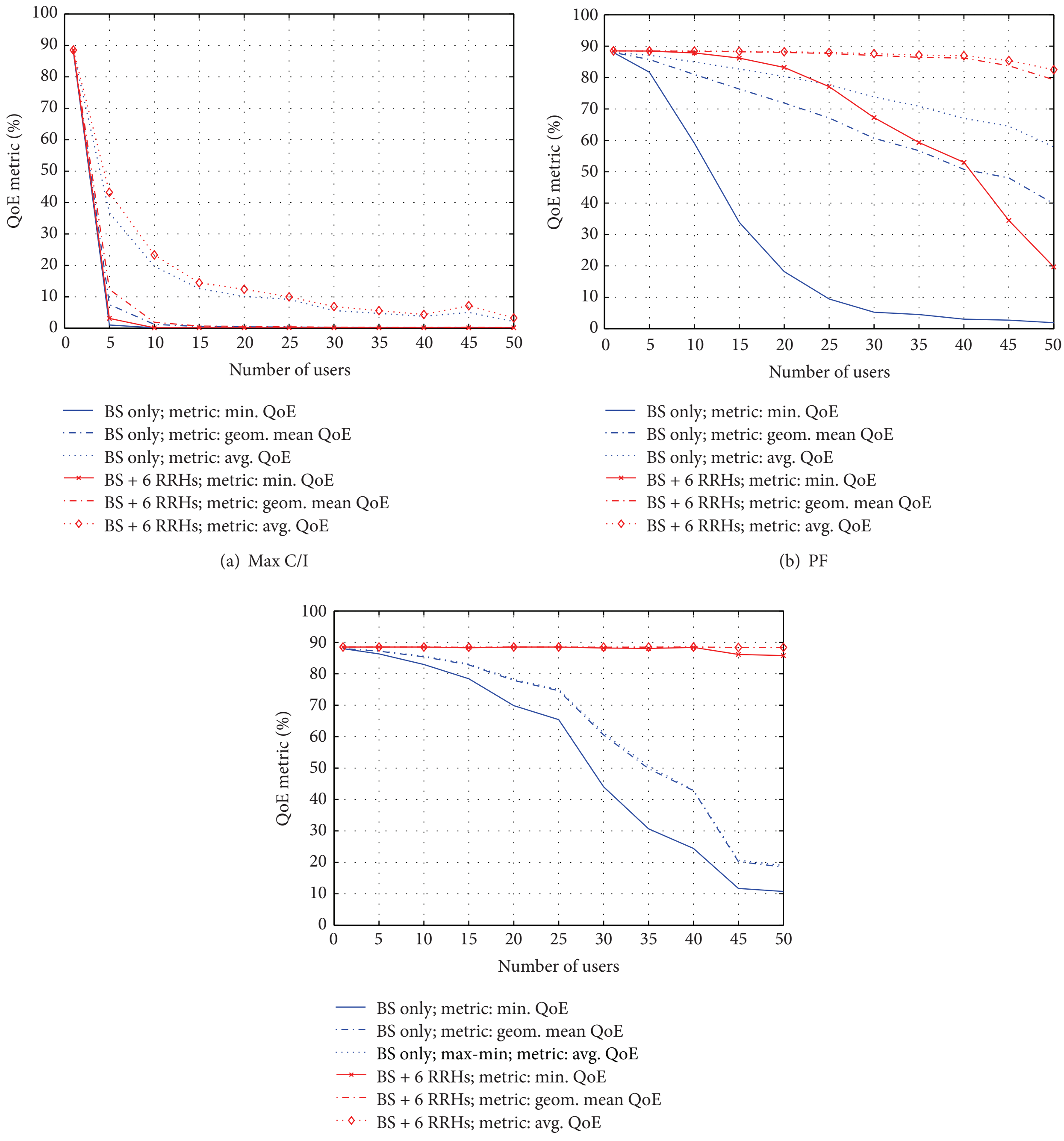
Therefore, in the next section, we consider the QoE network metrics and propose the use of DAS to enhance the performance and ensure more fairness to the users in the network.

6.3. Video QoE Results with DAS Deployment. In this section, we compare the performance of the two scenarios of Figure 2, using the QoE network metrics of Section 3 along with the RRM utilities described in Section 5. The DL results are presented in Figure 6, whereas the UL results are shown in Figure 7. The deployment of DAS leads to significant enhancements for all the investigated scenarios, except for the DL case with Max C/I scheduling, as shown in Figure 6(a), where the enhancement is minor. As explained in the previous section, Max C/I allocates the subcarriers and power available at the BS to users that are relatively close to the BS, which deprives users that are further away from wireless resources. This leads to very low QoE results for most users, especially due to error propagation in video sequences caused by lost frames.

A major difference between the results of Figure 6 and Figure 4(a) is that the Max-Min scheduler leads to the best results in the DL in the presence of DAS for all three network QoE metrics. In fact, Figure 6(c) shows that the performance for the three metrics is almost perfect: the horizontal curves indicate that the maximum QoE (88.5%) is reached for all users. The only reason for not having 100% QoE is due to source distortion caused by lossy compression of the video sequence and not to loss distortion due to packet losses over the wireless channels. The min QoE decreases only slightly when the number of users exceeds 40, as shown in Figure 6(c). This quasiideal performance was achieved with a DAS deployment using six RAHs throughout the cell.

In the UL, the best performance was achieved by PF scheduling, as shown in Figure 7 (particularly Figure 7(b)). The PF scheduler was also the best UL scheduler in the absence of DAS (Figure 4(b)), although DAS has led to a large performance enhancement for all network QoE metrics. This difference between the best DL (Max-Min) and UL (PF) schedulers is explained by the difference of power distribution between DL and UL. In the UL, the power is distributed over all users, where each user has an individual limited power. The deployment of DAS helps enhance the UL network QoE, but the limited transmit power still prevents worst case users from achieving near-optimal performance, which is captured by the min QoE network metric. On the other hand, the BS is the single source of power in the DL. The significantly larger transmit power at the BS provides better flexibility in power and subcarrier allocations and allows the enhancement of the QoE of the worst case users. Furthermore, the DAS deployment enhances the channel conditions for the users in the cell by providing transmit antennas closer to cell edge users. This allows the achievement of better individual QoE results with a lower transmit power for all users, including worst case users favored by Max-Min scheduling, which enhances the overall performance in the network.

Finally, it should be noted that the enhancements reached with DAS could be reached with other solutions. For example, dense heterogeneous network deployments, where small cell BSs are deployed in large numbers within the coverage area of large macrocell BSs, would lead to the same effects due to providing a complete (small) BS closer to the users instead of an RAH in a DAS deployment. Nevertheless, the DAS deployment, when possible, leads to a more cost effective solution. Furthermore, performance enhancements



(a) Max C/I

(b) PF

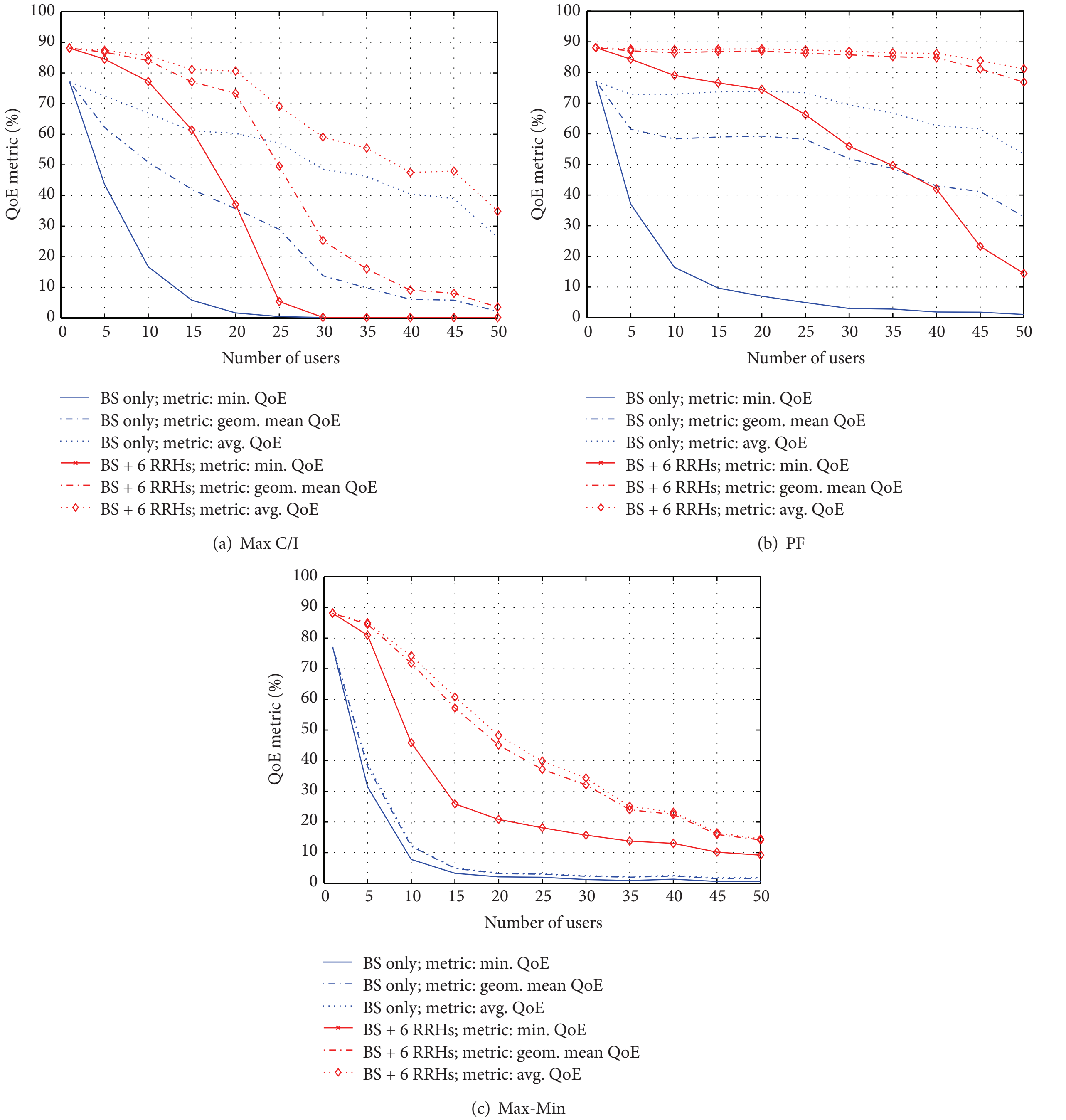
(c) Max-Min

Figure 6: DL QoE results with DAS.

for indoor users can be reached by deploying femtocell access points (FAPs). Obviously, FAP provides an indoor BS close to the end user, which allows the provision of high QoE by overcoming the penetration losses of the macrocell signal, coming from an outdoor BS located further away. The in-depth investigation of network QoE optimization in these specific scenarios is indeed an interesting topic for further research.

7. Conclusions

Network QoE metrics were proposed in order to capture the overall performance of radio resource management algorithms in terms of video quality perceived by the end users. Metrics corresponding to average, geometric mean, and minimum QoE in the network were measured when Max C/I, proportional fair, and Max-Min radio resource



(a) Max C/I

(b) PF

(c) Max-Min

Figure 7: UL QoE results with DAS.

management algorithms were implemented in the network. Both the uplink and downlink directions were studied. Furthermore, the use of distributed antenna systems to enhance the performance was considered. In the absence of distributed antennas, results showed that proportional fair scheduling maximizes the average network QoE both in the uplink and downlink. It also leads to more fairness in the uplink and in the downlink when the number of users increases by maximizing the geometric mean QoE. When distributed antennas are deployed, proportional fair scheduling was able to maximize the uplink performance in terms of network QoE, whereas Max-Min scheduling led to remarkably excellent results in the downlink with all the investigated network QoE metrics.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was made possible by NPRP Grant no. 4-347-2-127 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

[1] J.-S. Lee, F. de Simone, T. Ebrahimi, N. Ramzan, and E. Izquierdo, "Quality assessment of multidimensional video scalability," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 38–46, 2012.

[2] Video Quality Experts Group (VQEG), Final Report on the Validation of Objective Models of Video Quality Assessment, August 2003.

[3] H. Sohn, H. Yoo, W. de Neve, C. S. Kim, and Y. M. Ro, "Full-reference video quality metric for fully scalable and mobile SVC content," *IEEE Transactions on Broadcasting*, vol. 56, no. 3, pp. 269–280, 2010.

[4] Y.-F. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 286–298, 2011.

[5] H. G. Myung and D. J. Goodman, *Single Carrier FDMA: A New Air Interface for Long Term Evolution*, Wiley, 2008.

[6] T. Lunttila, J. Lindholm, K. Pajukoski, E. Tiirola, and A. Toskala, "EUTRAN uplink performance," in *Proceedings of the 2007 2nd International Symposium on Wireless Pervasive Computing (ISWPC '07)*, pp. 515–519, February 2007.

[7] "3rd Generation Partnership Project (3GPP)," 3GPP TS 36.213 3GPP TSG RAN Evolved Universal Terrestrial Radio Access (EUTRA) Physical layer procedures, version 11.4.0, Release 11, 2013.

[8] X. Zhao and X. Yang, "Downlink ergodic capacity analysis for wireless networks with cooperative distributed antenna systems," in *Proceedings of the 2012 IEEE International Conference on Communications (ICC '12)*, pp. 5911–5915, June 2012.

[9] E. Yaacoub and Z. Dawy, *Resource Allocation in Uplink OFDMA Wireless Systems: Optimal Solutions and Practical Implementations*, John Wiley and Sons/IEEE Press, New York, NY, USA, 2012.

[10] E. Yaacoub and O. Kubbar, "On the performance of distributed base stations in LTE public safety networks," in *Proceedings of the 8th IEEE International Wireless Communications and Mobile Computing Conference (IWCMC '12)*, pp. 927–932, August 2012.

[11] CPRI Specification V2.0, "Common Public Radio Interface (CPRI); interface specification," Tech. Rep., 2004.

[12] L. U. Choi, M. T. Ivrlač, E. Steinbach, and J. A. Nossek, "Analysis of distortion due to packet loss in streaming video transmission over wireless communication links," in *Proceedings of the IEEE International Conference on Image Processing 2005 (ICIP '05)*, pp. 189–192, September 2005.

[13] J.-S. Lee, F. de Simone, and T. Ebrahimi, "Subjective quality evaluation of foveated video coding using audio-visual focus of attention," *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1322–1331, 2011.

[14] P. Goudarzi, "Scalable video transmission over multi-hop wireless networks with enhanced quality of experience using swarm intelligence," *Signal Processing: Image Communication*, vol. 27, no. 7, pp. 722–736, 2012.

[15] X. Qiu, "On the performance of adaptive modulation in cellular systems," *IEEE Transactions on Communications*, vol. 47, no. 6, pp. 884–895, 1999.

[16] E. Yaacoub and N. Zorba, "Enhancing the quality of experience of video streaming in LTE networks using distributed antenna systems," in *Proceedings of the International Conference on Communications and Information Technology (ICCIT '13)*, Beirut, Lebanon, June 2013.

[17] J.-Y. le Boudec, *Rate Adaptation, Congestion Control and Fairness: A Tutorial*, EPFL, 2008.

[18] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks. Part I: theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 614–624, 2005.

[19] F. Carmichael, *A Guide to Game Theory*, Prentice Hall,, 2005.

[20] E. Yaacoub and Z. Dawy, "Achieving the Nash bargaining solution in OFDMA uplink using distributed scheduling with limited feedback," *International Journal of Electronics and Communications*, vol. 65, no. 4, pp. 320–330, 2011.

[21] 3rd Generation Partnership Project (3GPP), 3GPP TS 36.211 3GPP TSG RAN Evolved Universal Terrestrial Radio Access (EUTRA) Physical Channels and Modulation, version 11.4.0, Release 11, 2013.

[22] 3rd Generation Partnership Project (3GPP), 3GPP TR 25.814 3GPP TSG RAN Physical Layer Aspects for Evolved UTRA, v7.1.0, 2006.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 108673, 11 pages
http://dx.doi.org/10.1155/2014/108673

Research Article

Threshold Accepting Heuristic for Fair Flow Optimization in Wireless Mesh Networks

Jarosław Hurkała and Tomasz Śliwiński

Institute of Control & Computation Engineering, Warsaw University of Technology, Ulica Nowowiejska 15/19, 00-665 Warsaw, Poland

Correspondence should be addressed to Jarosław Hurkała; j.hurkala@elka.pw.edu.pl

Received 10 February 2014; Revised 27 April 2014; Accepted 7 May 2014; Published 25 May 2014

Academic Editor: Dritan Nace

Developing effective, fairness-preserving optimization algorithms is of considerable importance in systems which serve many users. In this paper we show the results of the threshold accepting procedure applied to extremely difficult problem of fair resource allocation in wireless mesh networks (WMN). The fairness is modeled by allowing preferences with regard to distribution of Internet traffic between network participants. As aggregation operator we utilize weighted ordered weighted averaging (WOWA). In the underlaying optimization problem, the physical medium properties cause strong interference among simultaneously operating node devices, leading to nonlinearities in the mixed-integer pricing subproblem. That is where the threshold accepting procedure is applied. We show that, the threshold accepting heuristic performs much better than the widely utilized simulated annealing algorithm.

1. Introduction

Wireless mesh network (WMN) is an organized cooperating group of network devices communicating with each other by means of wireless media. The nodes are organized in a mesh topology, where each wireless device not only sends and receives its own data but also serves as a relay for other nodes. Some of the nodes can be connected to cable network or mobile network and serve as Internet gateways. This way the whole mesh network constitutes a decentralized way of providing Internet access between attending participants.

This network type poses numerous advantages including setup cost, independence of the hardwired infrastructure, and flexibility. However, providing fair and efficient network management, including routing and scheduling, is not a straightforward task. The main source of difficulty lies in physical medium properties that cause strong interference among simultaneously operating devices. Additionally the link quality is a function of the distance and can be affected by obstacles present between the nodes. As a result the efficient network operation requires transmission scheduling, channel assignment, and transmission power determination.

Common objective of the optimization is maximization of the total throughput while retaining fairness in its distribution between participants.

In network optimization problems fairness is often accomplished using the lexicographic max-min (LMM) optimization. In the case of convex attainable set, this corresponds to the max-min fairness concept [1] which states that in the optimal solution it is impossible to increase any of the outcomes without the decreasing of some smaller (worse) ones [1–3]. In nonconvex case such strictly defined MMF solution may not exist while the LMM always exists and it covers the former if it exists (see [4] for wider discussion). However, LMM is a stiff approach that usually does not allow any other criteria, the overall efficiency (total throughput) in particular. Moreover, it requires sequential repeated optimization of the original problem. A recent survey on fairness oriented WMN optimization can be found in [5].

In the paper a more flexible approach based on the weighted ordered weighted averaging (WOWA) outcomes aggregation is proposed. It provides the consistent, reasonable, and fairness-preserving methodology of modeling various preferences (from the extreme pessimistic through neutral to extreme optimistic) with regard to distribution of Internet throughput between network participants. It is based on the OWA (ordered weighted aggregation) [6, 7] in which the preference weights are assigned to the ordered values (i.e., to the largest value, the second largest, and so on) rather than to the specific criteria. The WOWA extension allows using additional weights (called importance weights) assigned to the specific outcomes.

The OWA operator provides a parameterized family of aggregation operators, which include many of the well-known operators such as the maximum, the minimum, the k-order statistics, conditional minimax [8] known as conditional value at risk (CVaR) in the field of financial risk measurement, the median, and the arithmetic mean. The OWA satisfies the properties of strict monotonicity, impartiality, and, in the case of monotonic increasing weights, the property of equitability (satisfies the principle of transfers: equitable transfer of an arbitrary small amount from the larger outcome to a smaller outcome results in a more preferred achievement vector). Thus the OWA-based optimization generates the so-called equitably efficient solutions (cf. [9] for the formal axiomatic definition). According to [9, 10], equitable efficiency expresses the concept of fairness, in which all system entities have to be treated equally and in the stochastic problems equitability corresponds to the risk aversion [11]. Since its introduction, the OWA aggregation has been successfully applied to many fields of decision making [7, 12, 13]. When applying the OWA aggregation to multicriteria optimization, the weighting of the ordered outcome values causes the fact that the OWA optimization problem is nonlinear even for linear programming formulation of the original constraints and criteria. Yager has shown that the nature of the nonlinearity introduced by the ordering operation allows one to convert the OWA optimization into a mixed integer programming problem. It was shown [14] that the OWA optimization with monotonic weights can be formed as a standard linear program of higher dimension. Its significant extension introduced by Torra [15] incorporates importance weighting into the OWA operator forming the weighted OWA (WOWA) aggregation as a particular case of Choquet integral using a distorted probability as the measure. The WOWA averaging is defined by two weighting vectors: the preferential weights and the importance weights. It covers both the weighted means (with equal preferential weights) and the OWA averages (with equal importance weights) as special cases. Some of the example applications of importance weights (applied to specific outcomes) include definition of the size or importance of processes in a multiagent environment, setting scenario probability (if uniform objectives represent various possible values of the same uncertain outcome under several scenarios), or job priorities in scheduling problems. It was shown [16] that in the case of monotonic preferential weights WOWA aggregation can also be modeled by a mere linear extension of the original problem.

This paper extends and refines our initial work on the subject presented in [17]. The list- based threshold accepting heuristic applied to the pricing problem is described. The results are compared to the exact MIP formulation.

This paper is organized as follows. In the next section, we present the wireless mesh network optimization problem together with the outline of the solution approach. Next the WOWA aggregation operator is introduced. In the fourth section we deal with the pricing problem and show two alternative approximate solution algorithms. The last section presents the setup and the results of the computational experiments.

2. Flow Optimization in WMN

The WMN networking technology has been drawing an increased attention over the last years (see literature overview in [18, 19] and references therein). Due to complexity of the problem usually some sorts of simplifications are assumed. The problem considered in this paper can be stated as follows. There is given a WMN network with a number of nodes, routers and gateways. The nodes are interconnected wirelessly in compliance with all the physical constraints and requirements, including signal loss with increasing distance and interference occurring during simultaneous operation. Each node can be either sending or receiving data, but not both at the same time. There are a number of modulation and coding schemes (MCSs) used for communication between the nodes with different properties with regard to speed, maximum allowable interference, and the distance. Each MCS has its signal to interference plus noise ratio (SINR) requirement that must be fulfilled in order to successfully transmit data. Only one fixed transmitting power and single channel are assumed, but MCS can be dynamically allocated. The network model consists only of links for which at least one MCS can be applied, and this requirement reduces to the maximum allowable distance between the nodes.

Only downstream communication direction from gateways to routers is considered. For each router there is a single predefined path leading to a chosen gateway. The routers have elastic traffic demand, which means they can consume all the possible network capacity. The demands compete for network resources to get as much throughput as possible.

The objective is to maximize total throughput preserving fairness among competing demands.

The solution approach is based on the concept of compatible sets introduced in [20]. Compatible set consists of links that can operate at the same time within given interference model. The basic solution concept consists in linear approximation of the model and consecutive generation of the compatible sets improving current solution within the column generation schema. The approximation is needed if the time horizon is divided into fixed-length time slots; if not the solution is optimal.

Although we consider only a specific problem, the solution concepts involving application of WOWA operators can be utilized for many other variants of WMN problems including different capacity reservation models (see [19]).

2.1. Notation. Wireless mesh network topology is represented by a directed graph $\mathcal{N} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{G} \cup \mathcal{R}$ is the set of nodes from which we distinguish the set of gateways and set of mesh routers denoted, respectively, by $\mathcal{G}$ and $\mathcal{R}$, and $\mathcal{E}$ is the set of (radio) links.

The (potential) link between two nodes $v, w \in \mathcal{V}$ is modeled by a directed arc $e = (v, w) \in \mathcal{E}$, where $a(e) = v$ is the originating node that can transmit a signal of a given power P_{vw} to its terminating node $b(e) = w$. Additionally, we assume that if arc $e = (v, w) \in \mathcal{E}$ exists, then an opposite arc $e' = (w, v) \in \mathcal{E}$ also exists. Furthermore, the sets of outgoing and incoming arcs from/to node $v \in \mathcal{V}$ are denoted, respectively, by $\delta^+(v)$ and $\delta^-(v)$, while $\delta(v) = \delta^+(v) \cup \delta^-(v)$ is the set of all arcs incident to node v.

Nodes are transmitting using one of the available *modulation and coding schemes* (MCSs) denoted by $m \in \mathcal{M}$, where $\mathcal{M}$ is the set of all MCSs (to simplify the considerations, we assume that the set of available MCSs is $\mathcal{M}(e) = \mathcal{M}$, $e \in \mathcal{E}$). The (raw) data rate of transmission using MCS m is denoted by B^m.

The (radio) link $e = (v, w)$ can successfully transmit if the signal to noise ratio (SNR) for the arc e is greater than a certain threshold value denoted by γ^m for at least one MCS $m \in \mathcal{M}$:

$$\Gamma'_e = \frac{P_{vw}}{N} \geq \gamma^m, \tag{1}$$

where $N = 10^{-10.1}$ mW is the ambient noise power.

At any arbitrary time instance the transmission of other nodes can interfere with transmission on e. The corresponding signal to interference plus noise ratio (SINR) condition for successful transmission on e using MCS m reads as follows:

$$\Gamma_e = \frac{P_{vw}}{N + \sum_{a \in \mathcal{A} \setminus \{v\}} P_{aw}} \geq \gamma^m, \tag{2}$$

where $\mathcal{A} \subseteq \mathcal{V}$ is the set of active nodes which are transmitting at the same time.

Moreover, we assume that a node can either transmit or receive or be inactive; that is,

$$\begin{aligned} \mathcal{A} \cap \{b(\delta^+(v))\} \neq \emptyset \Longrightarrow \mathcal{A} \cap \{a(\delta^-(v))\} = \emptyset \quad v \in \mathcal{V}, \\ \mathcal{A} \cap \{a(\delta^-(v))\} \neq \emptyset \Longrightarrow \mathcal{A} \cap \{b(\delta^+(v))\} = \emptyset \quad v \in \mathcal{V}. \end{aligned} \tag{3}$$

Each router $r \in \mathcal{R}$ is connected with a selected gateway $g \in \mathcal{G}$ by a directed path p_d (i.e., a subset of links, $p_d \subseteq \mathcal{E}$) that is supposed to carry the entire downstream flow f_d from gateway g to router r (to simplify the formulations, we do not consider the upstream direction). The set of routers is considered as demands and denoted by $d \in \mathcal{D}$, where $\mathcal{D} = \mathcal{V} \setminus \mathcal{G}$. Let $\mathcal{P} = \{p_1, \ldots, p_D\}$ be the given set of paths between routers and gateways, where $D = |\mathcal{D}|$. For each link $e \in \mathcal{E}$, the set of all indices of paths in $\mathcal{P}$ that contain this link will be denoted by $\mathcal{Q}_e = \{d : e \in p_d, 1 \leq d \leq D\}$.

2.2. Compatible Sets. A compatible set (CS) is defined as a subset $\mathcal{E}_i$ of links $\mathcal{E}_i \subseteq \mathcal{E}$ together with a particular MCS m_e, $e \in \mathcal{E}_i$ that each link is using so that each link can be active simultaneously (i.e., transmit without generating too much interfering with other links). In other words, a compatible set is defined by $\mathcal{E}_i = \{(e, m) \in \mathcal{E} \times \mathcal{M} : y_e^m = 1\}$, where variables y_e^m form a feasible solution that satisfy (2) and (3).

2.2.1. Master Problem. Using the family of compatible sets denoted by $\mathcal{I}$, the formulation of the max-min flow optimization problem reads as follows:

$$\max f, \tag{4}$$

$$f \leq f_d \quad d \in \mathcal{D}, \tag{5}$$

$$\sum_{d \in \mathcal{Q}_e} f_d \leq c_e \quad e \in \mathcal{E}, \tag{6}$$

$$c_e = \sum_{i \in \mathcal{I}} B_{ei} z_i \quad e \in \mathcal{E}, \tag{7}$$

$$\sum_{i \in \mathcal{I}} z_i = T, \tag{8}$$

$$z \geq 0. \tag{9}$$

In the presented formulation, T is the time of network operation, B_{ei} is the (raw) data rate of a transmission using MCS $m \in \mathcal{M}$ allocated to link $e \in \mathcal{E}$ in compatible set $i \in \mathcal{I}$, that is, either B^m or 0, depending on whether link e is active or not in the compatible set i, and c_e is total amount of data that can be transmitted over link $e \in \mathcal{E}$ in a time interval T.

This formulation is a noncompact, continuous approximation of the MIP problem involving time slots (see [19]); continuous variables z_i define the number of time slots assigned to a compatible set within the time T.

Since $|\mathcal{I}|$ grows exponentially in the network size, the solution is to use the column generation technique [3, 21], where not all the columns of the constraints matrix are stored. Instead, only a subset of the variables (columns) that can be seen as an approximation (restriction) of the original problem is kept. The column generation algorithm iteratively modifies the subset of variables by introducing new variables in a way that improves the current optimal solution. At the end, the set contains all the variables necessary to construct the overall optimal solution which can use all possible columns. New columns are generated in the pricing problem.

2.2.2. Pricing Problem. The pricing problem we consider corresponds to a WMN system in which there are multiple MCSs available and each node can use different MCS in different compatible set. The following formulation is referred to as dynamic allocation of MCSs to nodes:

$$\max \sum_{e \in \mathcal{E}} \pi_e^* B_e, \tag{10}$$

$$X_v = \sum_{m \in \mathcal{M}} x_v^m \quad v \in \mathcal{V}, \tag{11}$$

$$Y_e = \sum_{m \in \mathcal{M}} y_e^m \quad e \in \mathcal{E}, \tag{12}$$

$$\sum_{e\in\delta(v)} Y_e \le 1 \quad v \in \mathcal{V}, \tag{13}$$

$$\sum_{e\in\delta^+(v)} y_e^m = x_v^m \quad v \in \mathcal{V},\ m \in \mathcal{M}, \tag{14}$$

$$u_{ev}^m \ge y_e^m + X_v - 1 \quad v \in \mathcal{V},\ e \in \mathcal{E},\ m \in \mathcal{M}, \tag{15}$$

$$u_{ev}^m \le y_e^m, \quad u_{ev}^m \le X_v \quad v \in \mathcal{V},\ e \in \mathcal{E},\ m \in \mathcal{M}, \tag{16}$$

$$\begin{aligned} N y_e^m + \sum_{v\in\mathcal{V}\setminus\{a(e)\}} P_{vb(e)} u_{ev}^m \le \frac{1}{\gamma^m} P_{a(e)b(e)} y_e^m \\ e \in \mathcal{E},\ m \in \mathcal{M}, \end{aligned} \tag{17}$$

$$B_e = \sum_{m\in\mathcal{M}} B^m y_e^m \quad e \in \mathcal{E}, \tag{18}$$

$$y_e^m \in \{0, 1\}. \tag{19}$$

In this formulation, π_e, $e \in \mathcal{E}$, are the current optimal dual variables associated with constraints (6). At each iteration we are interested in generating CS for which the reduced price $\pi \cdot B$ has the biggest and positive value, as we can expect this will improve the current optimal solution as much as possible. The x_v^m is a binary variable indicating whether node v transmits using MCS m, X_v is a binary variable indicating whether node v transmits, y_e^m is a binary variable indicating whether link e is scheduled to be active with the MCS m, Y_e is a binary variable indicating whether link e is active, and B_e is the (raw) data rate of a transmission allocated to link e. Notice that, in this formulation, X_v and Y_e are auxiliary variables and thus either they can be eliminated or their binarity can be skipped. Moreover, observe that applying (2) directly to our model would result in a bilinear constraint. Hence, we have introduced additional (continuous) variables u_{ev}^m to make the constraint (17) linear. This is achieved by adding the constraints (15)-(16); that is, $u_{ev}^m = 1$ if both y_e^m and X_v are equal to 1, and 0, otherwise.

Each node $v \in \mathcal{V}$ and each link $e \in \mathcal{E}$ can use at most one MCS $m \in \mathcal{M}$ in the compatible set (11)-(12). At most one link $e \in \delta(v)$ incident to node v can be active (13) and exactly one link $e \in \delta^+(v)$ outgoing from node v is active and uses MCS m (14), provided the node is active and uses this MCS in the compatible set. The constraints (15)–(17) assure admissible SINR for link e using MCS m in the compatible set. The (raw) data rate B_e of link e in the compatible set is found by (18).

3. Fair Aggregation Operators

As stated before the basic operator used to preserve fairness among outcomes is lexicographic max-min (LMM) which is equivalent to MMF for the linear problems. In such a case it is possible to carry out the MMF procedure based on simple algorithm that in each step uses the dual information to determine the outcomes that are blocked at their highest values possible. In the following steps, only the outcomes are optimized that have not been blocked before (for details see [19]).

On the other hand, in the WOWA aggregation, the original problem is extended by auxiliary constraints and solved in a single step. Let us first introduce the formalization of the OWA operator. In the OWA aggregation of outcomes $\mathbf{y} = (y_1, \ldots, y_n)$ preferential weights $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ are assigned to the ordered values rather than to the specific criteria:

$$A_w = \sum_{i=1}^{n} w_i \theta_i(\mathbf{y}), \tag{20}$$

where $(\theta_1(\mathbf{y}), \theta_2(\mathbf{y}), \ldots, \theta_n(\mathbf{y})) = \Theta(\mathbf{y})$ is the ordering map $R^n \to R^n$ with $\theta_1(\mathbf{y}) \le \theta_2(\mathbf{y}) \le \cdots \le \theta_n(\mathbf{y})$ and there exists a permutation τ of set I such that $\theta_i(\mathbf{y}) = y_{\tau(i)}$ for $i = 1, 2, \ldots, n$.

If the weights are monotonic, $w_1 > w_2 > \cdots > w_{n-1} > w_n$, the OWA aggregation has the property of equitability [14], which guarantees that an equitable transfer of an arbitrarily small amount from the larger outcome to a smaller outcome results in more preferred achievement vector. Every solution maximizing the OWA function is then an equitably efficient solution to the original multiple criteria problem. Moreover, for linear multiple criteria problems every equitably efficient solution can be found as an optimal solution to the OWA aggregation with appropriate weights.

For the maximization problem the OWA objective aggregation can be formulated as linear extension of the original problem, as follows. Let us apply linear cumulative map to the ordered achievement vectors $\Theta(\mathbf{y})$:

$$\overline{\theta}_k(\mathbf{y}) = \sum_{i=1}^{k} \theta_i(\mathbf{y}) \quad k = 1, 2, \ldots n. \tag{21}$$

As stated in [14], for any given vector $\mathbf{y} \in R^n$, the cumulated ordered coefficient $\overline{\theta}_k(\mathbf{y})$ can be found as the optimal value of the following LP problem:

$$\begin{aligned} \overline{\theta}_k(\mathbf{y}) = \max \quad & k t_k - \sum_{i=1}^{n} h_{ki}, \\ \text{s.t.} \quad & t_k - y_i \le h_{ki}, \quad h_{ki} \ge 0 \\ & i = 1, 2, \ldots, n. \end{aligned} \tag{22}$$

The ordered outcomes can be expressed as differences $\theta_i(\mathbf{y}) = \overline{\theta}_i(\mathbf{y}) - \overline{\theta}_{i-1}(\mathbf{y})$ for $i = 2, \ldots, n$ and $\theta_1(\mathbf{y}) = \overline{\theta}_1(\mathbf{y})$. Hence, the maximization of the OWA operator (20) with weights w_i can be expressed in the form

$$\max \left\{ \sum_{i=1}^{n} w_i' \overline{\theta}_i(\mathbf{y}) : \mathbf{y} \in Y \right\}, \tag{23}$$

where coefficients w_i' are defined as $w_n' = w_n$ and $w_i' = w_i - w_{i+1}$ for $i = 1, 2, \ldots, n-1$ and Y is the feasible set of outcome vectors $\mathbf{y}$. If the original weights w_i are strictly decreasing, then $w_i' > 0$ for $i = 1, 2, \ldots, n$.

For the WMN flow optimization problem (4)–(9) the final OWA aggregation of the outcomes f_d for all demands/routers can be stated as the following LP model:

$$\max \sum_{k=1}^{D} k w_k' t_k - \sum_{k=1}^{D} \sum_{d=1}^{D} w_k' h_{dk}, \tag{24}$$

subject to

$$h_{dk} \geq t_k - f_d, \quad h_{dk} \geq 0 \quad d, k = 1, 2, \ldots, D$$
$$\mathbf{f} \in F, \tag{25}$$

where $\mathbf{f} = [f_d]_{d\in\mathcal{D}}$ and F is a feasible set of flows/throughputs defined by (6)–(9).

The OWA aggregation (20) allows modelling various aggregation functions from the minimum ($w_1 = 1, w_i = 0$ for $2, \ldots, n$), through the arithmetic mean ($w_i = 1/n$ for $i = 1, \ldots, n$), to the maximum ($w_n = 1, w_i = 0$ for $i = 1, \ldots, n-1$). However, it is not possible to express the weighted mean. Due to the property of impartiality and neutrality with respect to the individual attributes the OWA aggregation does not allow representing any importance weights allocated to the specific attributes.

The WOWA aggregation is a generalization of the OWA that allows assigning importance weights to specific criteria [22]. In the case of WMN, the importance weights could express the number of end-users hidden behind each router. For example, the importance weight of the router with 5 directly connected users should be 5 times greater than the importance weight of the router with only a single directly connected user.

Let $\mathbf{p} = (p_1, \ldots, p_n)$ be an n-dimensional vector of importance weights such that $p_i \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} p_i = 1$. The corresponding weighted OWA aggregation of vector $\mathbf{y}$ is defined [15] as follows:

$$A_{w,p} = \sum_{i=1}^{n} \omega_i \theta_i(\mathbf{y}), \tag{26}$$

with

$$\omega_i = w^* \left(\sum_{k \leq i} p_{\tau(k)} \right) - w^* \left(\sum_{k < i} p_{\tau(k)} \right), \tag{27}$$

where w^* is increasing function interpolating points $(i/n, \sum_{k\leq i} w_k)$ together with the point $(0, 0)$ and τ representing the ordering permutation for $\mathbf{y}$ (i.e., $y_{\tau(i)} = \theta(\mathbf{y})$). Moreover, function w^* is required to be a straight line when the points can be interpolated in this way. Due to this requirement, the WOWA aggregation covers the standard weighted mean with weights p_i as a special case of equal preference weights ($w_i = 1/n$ for $i = 1, \ldots, n$). Actually, the WOWA operator is a particular case of the Choquet integral using a distorted probability as the measure [23].

Note that such WOWA definition allows us for a clear interpretation of importance weights as the agent (demand) repetitions [24]. Splitting an agent into two agents does not cause any change of the final distribution of outcomes. For theoretical considerations one may assume that the problem can be transformed (disaggregated) to the unweighted one (that means all the agent importance weights are equal to 1); see [22, 25] and examples therein. Thus, the WOWA aggregation with increasing preferential weights is equitable since equally important unit of a smaller outcome is considered with a larger weight.

As shown in [22], maximization of an equitable WOWA aggregation with decreasing preferential weights $w_1 > w_2 > \cdots > w_n$ may also be implemented as the LP expansion of the original problem. In the case of the WMN flow optimization problem (6)–(9), this can be stated as follows:

$$\max \sum_{k=1}^{D} w'_k \left[\frac{k}{n} t_k - \sum_{d=1}^{D} p_d h_{dk} \right], \tag{28}$$

subject to

$$h_{dk} \geq t_k - f_d, \quad h_{dk} \geq 0 \quad d, k = 1, 2, \ldots, D, \tag{29}$$

$$\mathbf{f} \in F. \tag{30}$$

If the importance weights are equal to $p_d = 1/D$, the model reduces to the OWA aggregation.

A special case of the generalized WOWA aggregation is defined for single breakpoint and corresponds to optimization of the predefined quantile of the worst outcomes and in finance is known as the CVaR (conditional value at risk). It can be computed as a standard linear extension of the original problem [22]:

$$\max t - \frac{1}{\beta} \sum_{d=1}^{D} p_d h_d \tag{31}$$

subject to

$$h_d \geq t - f_d, \quad h_d \geq 0 \quad d = 1, \ldots, D \tag{32}$$

$$\mathbf{f} \in F. \tag{33}$$

4. Algorithms

4.1. List-Based Threshold Accepting. List-based threshold accepting algorithm (LBTA) is an extension of threshold accepting metaheuristic, which belongs to the randomized search class of algorithms. This rather unknown heuristic has been successfully applied to many difficult problems [26–30]. Since the problem of fair resource allocation in wireless mesh networks is extremely challenging, we have decided to try this underappreciated algorithm.

The search trajectory of LBTA crosses the solution space by moving from one solution to a random neighbor of that solution and so on. Unlike the greedy local search methods which consist of choosing a better solution from the neighborhood of the current solution until such can be found (hill climbing), the threshold accepting allows choosing a worse candidate solution based on a threshold value. In the general concept of the threshold accepting algorithm it is assumed that a set of decreasing threshold values is given before the computation or an initial threshold value and a decrease schedule are specified. The rate at which the values decrease controls the trade-off between diversification (associated with large threshold values) and intensification (small threshold values) of the search. It is immensely difficult to predict how the algorithm will behave when a certain decrease rate is applied for a given problem without running the actual

```
Require: Initial solution s_1, list size S, set of move operators m ∈ M
i ← 0
while i < N do
    m ← random(M)
    s_2 ← m(s_1)
    if C(s_1) ≤ C(s_2) then
        Δ ← (C(s_2) − C(s_1))/C(s_1)
        list ← list ∪ {Δ}
        i ← i + 1
    else
        s_1 ← s_2
    end if
end while
return list
```

Algorithm 1: Creating the list of threshold values.

computation. It is also very common that the algorithm with the same parameters works better for some problem instances and significantly worse for others. These reflections led to the list-based threshold accepting branch of threshold accepting metaheuristic.

In the list-based threshold accepting approach, instead of a predefined set of values, a list is dynamically created during a presolve phase of the algorithm. The list, which in a way contains knowledge about the search space of the underlying problem, is then used to solve it.

4.1.1. Creating the List of Threshold Values. The first phase of the algorithm consists of gathering information about the search space of the problem that is to be solved. From an initial solution a neighbor solution is created using a move function (perturbation operator) chosen at random from a predefined set of functions. If the candidate solution is better than the current one, it is accepted and becomes the current solution. Otherwise, a threshold value is calculated as a relative change between the two solutions,

$$\Delta = \frac{C(s_2) - C(s_1)}{C(s_1)}, \quad (34)$$

and added to the list, where $C(s_i)$ is the objective function value of the solution $s_i \in S$ and S is a set of all feasible solutions. For this formula to work, it is silently assumed that $C : S \rightarrow \mathbb{R}_+ \cup \{0\}$. This procedure is repeated until the specified size of the list is reached. For the algorithm overview see Algorithm 1.

4.1.2. Optimization Procedure. The second phase of the algorithm is the main optimization routine, in which a solution to the problem is found. The algorithm itself is very similar to that of the previous phase. We start from an initial solution, create new solution from the neighborhood of current one using one of the move functions, and compare both solutions. If the candidate solution is better, it becomes the current one. Otherwise, a relative change is calculated. To this point algorithms in both phases are identical. The difference in the optimization procedure is that we compare the threshold value with the largest value from the list. If the new threshold value is larger, then the new solution is discarded. Otherwise, the new threshold value replaces the value from the list, and the candidate solution is accepted for the next iteration. The best solution found during the optimization process is considered final.

The list-based threshold accepting algorithm also incorporates early termination mechanism: after a (specified) number of candidate solutions are subsequently discarded, the optimization is stopped and the best solution found so far is returned.

The optimization procedure of the list-based threshold accepting algorithm is shown in Algorithm 2.

4.2. Simulated Annealing. Simulated annealing (SA) was first introduced by Kirkpatrick et al. [31], while Černý [32] pointed out the analogy between the annealing process of solids and solving combinatorial problems. Applications of the SA algorithm to optimization problems in various fields have been studied [33–36] and to the WMN problem, as well [17].

The optimization process of the simulated annealing algorithm can be described in the following steps. At the start, an initial solution is required. Then, repeatedly, a candidate solution is randomly chosen from the neighborhood of the current solution. If the candidate solution is the same or better than the current one, it is accepted and replaces the current solution. Even if the generated solution is worse than the current one, it still has a chance to be accepted with the so-called acceptance probability. This probability is a function of difference between objective value of the current and the candidate solution and depends on a control parameter taken from the thermodynamics, called temperature (τ). After a number of iterations, the temperature is decreased by a reduce factor (α), and the process continues as described above. The optimization is stopped either after a maximum number of iterations or when a minimum temperature is reached. The best solution found during the annealing process is considered final.

Require: Initial solution s_1, thresholds list L, set of move operators $m \in M$
$i \leftarrow 0$
$s^* \leftarrow s_1$
while $i \le N$ **do**
  $m \leftarrow \mathrm{random}(M)$
  $s_2 \leftarrow m(s_1)$
  $i \leftarrow i + 1$
  if $C(s_2) \le C(s_1)$ **then**
    if $C(s_2) \le C(s^*)$ **then**
      $s^* \leftarrow s_2$
    end if
    $s_1 \leftarrow s_2$
    $i = 0$
  else
    $\Delta_{\text{new}} \leftarrow (C(s_2) - C(s_1))/C(s_1)$
    if $\Delta_{\text{new}} < \max(\text{list})$ **then**
      $\text{list} \leftarrow \text{list} \setminus \{\max(\text{list})\}$
      $\text{list} \leftarrow \text{list} \cup \{\Delta_{\text{new}}\}$
      $s_1 \leftarrow s_2$
      $i = 0$
    end if
  end if
end while
return s^*

Algorithm 2: LBTA optimization procedure.

Table 1: Simulated annealing parameters.

Parameter	Description	Value
α	Reduce factor	$1 - \dfrac{7}{N}$
τ^0	Initial temperature	0.99
δ^0	Minimal difference between solutions	0.01
p^0	Initial acceptance probability	1
N	Number of SA iterations	300000
N_{const}	Number of iterations at constant temperature	10

For the algorithm overview see Algorithm 3, and for the overview of the SA parameters see Table 1.

4.3. Neighborhood Function. The most problem-specific mechanism of both the SA and the LBTA algorithm that always needs a different approach and implementation is the procedure of generating a candidate solution from the neighborhood of the current one, which is called a perturbation scheme, transition operation/operator, or a move function. Although there are many ways to accomplish this task, we have examined the following operators.

(1) Deactivate a node at random.
(2) Activate a random node and select the MCS with the smallest raw rate.
(3) Switch MCS to one with higher raw rate.
(4) Switch MCS to one with lower raw rate.
(5) Switch MCS to a random one.

Require: Initial solution s_1
$s^* \leftarrow s_1$
for $i = 1$ **to** N **do**
  for $t = 1$ **to** N_{const} **do**
    $s_2 \leftarrow \mathit{perturbate}(s_1)$
    $\delta \leftarrow C(s_2) - C(s_1)$
    if $\delta \le 0$ or $e^{-\delta/k\tau} > \mathrm{random}(0, 1)$ **then**
      $s_1 \leftarrow s_2$
    end if
    if $C(s_2) < C(s^*)$ **then**
      $s^* \leftarrow s_2$
    end if
  end for
  $\tau \leftarrow \tau * \alpha$
end for
return s^*

Algorithm 3: Simulated annealing.

In order to generate a new solution, the LBTA algorithm applies one of the aforementioned operators chosen at random to the current solution. SA on the other hand uses during the whole optimization procedure only one, compound operator, a combination of operators 1, 2, and 5 so that a transition from initial to any feasible solution is possible (see Algorithm 4).

4.4. Implementation Details

4.4.1. Zero Elements. In the first phase of the list-based threshold accepting algorithm the list is populated with

Require: Current solution compatible set CS
Ensure: $\mathrm{CS}' = \mathrm{neighbor}(\mathrm{CS})$
(1) Choose at random $v \in \mathscr{V}$ and $e \in \delta^{+}(v)$ that satisfy (11)–(14).
(2) **if** $v \in \mathscr{A}$ **then**
(3) **if** $\mathrm{random}(0,1) < 1/|\mathscr{M}(e)|$ **then**
(4) $\mathscr{A} \leftarrow \mathscr{A} \setminus \{v\}$ [deactivate node]
(5) **else**
(6) $m \leftarrow \mathrm{random}(\mathscr{M}(e) \setminus \{m\})$ [switch MCS]
(7) **end if**
(8) **else**
(9) $\mathscr{A} \leftarrow \mathscr{A} \cup \{v\}$ [activate node]
(10) $m \leftarrow \mathrm{random}(\mathscr{M}(e))$ [select MCS]
(11) **end if**

Algorithm 4: SA compatible set perturbation scheme.

values of relative change between two solutions $\Delta \geq 0$. After careful consideration, we believe that including zeros in the list is a misconception. In the actual optimization procedure, that is, the second phase, the threshold value is computed only if the new solution is worse than the current one, which means that the calculated relative change will always have a positive value ($\Delta_{\mathrm{new}} > 0$). The new threshold value is compared with the largest value from the list ($T_{h\max}$). Thus, we can distinguish three cases.

(1) $T_{h\max} = 0$: since thresholds are nonnegative from definition, in this case, the list contains all zero elements and it will not change throughout the whole procedure ($T_{h\max}$ is constant). Comparing a positive threshold value Δ_{new} against zero yields in discarding the candidate solution. The conclusions are as follows:

   (a) it does not matter how many zeros are in the list; the effective size of the list is equal to one;

   (b) the algorithm is reduced to hill climbing algorithm that accepts candidate solutions which are at least as good as the current one.

(2) $T_{h\max} > 0$ and $\Delta_{\mathrm{new}} < T_{h\max}$: the largest (positive) threshold value from the list $T_{h\max}$ is replaced by a smaller (positive) threshold value Δ_{new}. The number of zero elements in the list remains the same throughout the whole procedure and therefore is completely irrelevant to the optimization process. The effective list size is equal to the number of positive elements.

(3) $T_{h\max} > 0$ and $\Delta_{\mathrm{new}} \geq T_{h\max}$: the new solution is discarded and the list remains unchanged.

The main idea behind the list is to control the diversification and intensification of the search process. In the early stage of the search, the algorithm should allow covering as much solution space as possible, which means that the thresholds in the list are expected to be large enough to make that happen. In the middle stage, the algorithm should slowly stop fostering the diversification and begin to foster the intensification of the search. In the end stage, the intensification should be the strongest; that is, the list is supposed to contain smaller and smaller threshold values, which induces the discarding of worse solution candidates. As a consequence of that, the algorithm is converging to a local or possibly even a global optimum.

4.4.2. Stopping Criterion. Even though equipped with an early termination mechanism, the LBTA algorithm does not have a solution space independent stopping criterion. If the number of subsequently discarded worse solutions is set too high, the algorithm will run for an unacceptable long time (it has been observed during preliminary tests). Hence, we propose using a global counter of iterations so that when a limit is reached, the algorithm terminates gracefully.

5. Numerical Experiments

The problem defined by the constraints (6)–(9) with the network flows f_d as the optimization criteria was optimized with different aggregation operators: max-min, lexicographic max-min (LMM), CVaR (31)-(32), and WOWA (28)-(29). For the pricing problem (10)–(18) we applied the two approximate methods: the list-based threshold accepting (LBTA) algorithm and the simulated annealing (SA) algorithm and compared them to the exact MIP approach solved using the CPLEX 12.1 optimization package.

The numerical experiments were performed on a number of randomly generated problem instances of different sizes.

The algorithm of generating network topology instances can be described as follows. A grid of length 25 m of 30×30 points is created. Each of the grid points can be chosen to be a mesh router or a mesh gateway. First, the location of each gateway $g \in \mathscr{G}$ is chosen at random. Then, for each router $r \in \mathscr{R}$, a location is chosen at random that satisfies condition (1) for at least one link $e = (g, r)$, $g \in \mathscr{G}$, and MCS $m \in \mathscr{M}$. This condition is equivalent to $d_{gr} \leq d^{m}$, where d_{gr} is the distance between gateway g and router r and d^{m} is the maximum distance for selected MCS m. Finally, paths rooted in the gateways are established by iteratively connecting the neighboring routers that are reachable with the highest link rate and, if possible, with the lowest hop count. The specific data for different MCSs are presented in Table 2.

Table 2: IEEE 802.11a MCS, FER 61%, and 1500-byte payload.

MCS m	Raw rate B^m (Mbps)	SINR threshold $\hat{\gamma}^m$ (dB)	Maximum link length d^m (m)
BPSK 1/2	6	3.5	273.5
BPSK 3/4	9	6.5	230.0
QPSK 1/2	12	6.6	228.0
QPSK 3/4	18	9.5	193.7
16-QAM 1/2	24	12.8	160.2
16-QAM 3/4	36	16.2	131.7
64-QAM 2/3	48	20.3	103.8
64-QAM 3/4	54	22.1	93.5

Table 3: Computing times (s).

Problem size $\|\mathscr{D}\|, \|\mathscr{E}\|$	$\|\mathscr{G}\|$	Algorithm	Aggregation operator Max-Min	LMM	CVaR	WOWA
50	8	SA	160.8	214.6	116.2	85.2
10	2	SA	9.0	10.1	8.1	5.7
10	4	SA	8.9	10.6	7.9	4.7
10	8	SA	8.4	11.6	7.4	4.3
20	2	SA	34.0	39.0	27.0	22.8
20	4	SA	31.4	40.5	24.8	20.0
20	8	SA	29.8	41.0	27.1	16.4
50	2	SA	205.5	235.8	162.3	145.8
50	4	SA	157.0	188.6	127.3	109.2
10	2	TA	1.0	1.3	1.0	0.9
10	4	TA	1.0	1.3	0.9	0.7
10	8	TA	0.9	1.5	0.9	0.7
20	2	TA	4.4	5.1	3.5	3.9
20	4	TA	4.0	6.3	3.7	3.2
20	8	TA	3.9	6.2	4.0	2.6
50	2	TA	17.1	22.2	16.4	23.1
50	4	TA	11.8	21.3	12.6	16.0
50	8	TA	15.2	21.2	13.3	16.0
10	2	MIP	1.5	1.8	1.5	1.3
10	4	MIP	1.2	1.8	0.9	0.7
10	8	MIP	1.1	2.0	0.9	0.9
20	2	MIP	66.9	86.4	63.0	82.7
20	4	MIP	48.2	79.7	42.0	61.7
20	8	MIP	63.8	87.7	62.4	64.5
50	2	MIP	—	—	—	—
50	4	MIP	—	—	—	—
50	8	MIP	—	—	—	—

Although preferential weights determination is an important issue in the theory of ordered weighted averaging [37–39], for the performance check simple generation methodology has been chosen. All the weights, except two, are strictly decreasing numbers with the step 0.1, while the two selected weights ($k = \lfloor n/3 \rfloor$ and $k = \lfloor 2n/3 \rfloor$) differ from the previous ones by 0.5. The importance weights were generated as random values uniformly distributed in the range $[1, 2]$ and then normalized.

For the LBTA algorithm we used the list size of 50000 elements. This value was chosen based on our preliminary tests for selected problem sizes.

To better compare relative performance of the LBTA and SA algorithms, the only stopping criterion for single run was reaching exactly 300000 iterations, the same for all computations and problem sizes. This way we could compare the speed and the convergence per iteration.

All the experiments were performed on the Intel Core i7 3.4 GHz microprocessor using CPLEX 12.1 optimization library for the linear master problem. The results are the average of 10 randomly generated problems of a given size. Computing times are presented in Table 3, optimal objective value in Table 4, and the total number of the columns (compatible sets) generated (with either LBTA or SA) in Table 5. In all tables the test cases for which the timeout of 600 s occurred were marked with a "—" sign.

A note is required on the LMM optimal problem values (Table 4). Here the objective value of the last step of the LMM algorithm is given. That means that if in the previous steps only suboptimal values were reached, in the last step a better (greater) value than in the exact algorithm is possible. That is why the optimal objective values for the LMM should be treated only as a hint to the performance of the algorithm and not as an algorithm absolute quality measure.

The most noticeable advantage of LBTA algorithm over the SA algorithm when applied to the WMN problem is the computing time; in many cases the LBTA is faster than SA by an order of magnitude. The reason for such a good behavior lays not only in the computing speed of LBTA but also in the quality of the results because this affects the number of the generated columns (compatible sets). One can also notice generally better quality of the results, particularly for bigger problems, also compared to the exact MIP model.

We can also observe that the computation time increases with the number of routers. This is obvious and can easily be explained; as the network grows, the number of variables increases; hence, more compatible sets need to be computed. More interesting thing is that, for the same number of routers, the computation time is the highest for the smallest number of gateways. This can be explained by the fact that when the number of gateways is small, many paths between routers and gateways share the same link, which makes finding (feasible) compatible sets more difficult. This property has especially significant impact on the WOWA aggregation operator, for which the computation time is equal to or decreases with the number of gateways as long as the number of routers is fixed.

6. Conclusion

Effective, general purpose techniques are of considerable importance in many optimization areas. We have shown

Table 4: Optimal problem values (maximization).

Problem size $\|\mathscr{D}\|, \|\mathscr{E}\|$	$\|\mathscr{G}\|$	Algorithm	Aggregation operator Max-Min	LMM	CVaR	WOWA
10	2	SA	2.7	3.2	2.7	29.7
10	4		3.2	8.4	3.2	41.5
10	8		3.6	8.4	3.6	54.0
20	2		1.3	3.4	1.3	39.1
20	4		1.7	9.1	1.7	60.6
20	8		2.2	7.7	2.5	96.0
50	2		0.6	2.6	0.6	90.6
50	4		0.8	3.0	0.8	136.6
50	8		1.2	7.1	1.2	213.9
10	2	TA	2.7	3.2	2.7	29.7
10	4		3.2	8.0	3.2	41.4
10	8		3.6	10.2	3.6	54.0
20	2		1.3	4.2	1.3	38.9
20	4		1.7	14.1	1.7	60.0
20	8		2.0	10.2	2.4	95.0
50	2		0.6	4.5	0.6	93.2
50	4		0.6	6.0	0.7	139.4
50	8		1.2	7.0	1.2	215.9
10	2	MIP	2.8	3.3	2.8	30.2
10	4		3.3	7.9	3.3	42.6
10	8		3.8	7.9	3.8	54.8
20	2		1.4	3.2	1.4	41.2
20	4		1.8	11.2	1.8	63.0
20	8		2.6	5.5	2.7	99.4
50	2		—	—	—	—
50	4		—	—	—	—
50	8		—	—	—	—

Table 5: Number of compatible sets generated.

Problem size $\|\mathscr{D}\|, \|\mathscr{E}\|$	$\|\mathscr{G}\|$	Algorithm	Aggregation operator Max-Min	LMM	CVaR	WOWA
10	2	SA	15.5	17.6	14.3	10.6
10	4		16.5	19.9	15.6	10.4
10	8		16.1	22.8	14.8	9.5
20	2		39.2	47.9	35.8	32.2
20	4		40.9	56.7	35.5	32.5
20	8		43.9	63.3	42.0	28.9
50	2		132.9	161.6	124.6	118.9
50	4		119.0	151.3	109.4	100.8
50	8		136.8	198.8	112.8	90.1
10	2	TA	15.5	18.0	14.6	10.8
10	4		16.2	19.3	14.5	10.3
10	8		14.4	21.5	13.9	9.6
20	2		39.5	44.9	32.5	32.4
20	4		39.6	60.0	36.5	28.8
20	8		35.3	56.0	37.5	27.1
50	2		102.5	133.5	99.5	99.1
50	4		77.5	135.5	80.6	82.0
50	8		104.1	137.2	90.2	83.2
10	2	MIP	19.1	21.0	16.5	12.0
10	4		21.1	24.2	17.0	10.7
10	8		20.1	25.4	16.5	9.9
20	2		52.0	58.2	41.1	37.8
20	4		53.1	64.6	40.3	34.3
20	8		60.3	69.5	49.6	30.6
50	2		—	—	—	—
50	4		—	—	—	—
50	8		—	—	—	—

that one of the most promising algorithms is the list-based threshold accepting metaheuristic. When applied to the pricing problem of WMN optimization it gives a tremendous advantage over the classic and widely utilized simulated annealing algorithm. We have also shown that OWA-based advanced aggregation operators applied to the flow optimization in wireless mesh networks can be effectively modeled and solved when compared to the traditional lexicographic max-min operators.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The research was partially supported by the National Science Centre, Poland, under Grant no. 2011/01/B/ST7/02967 "Integer programming models for joint optimization of link capacity assignment, transmission scheduling, and routing in fair multicommodity flow networks."

References

[1] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, Englewood Cliffs, NJ, USA, 1987.

[2] J. Jaffe, "Bottleneck flow control," *IEEE Transactions on Communications*, vol. 7, pp. 207–237.

[3] M. Pióro and D. Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, San Francisco Calif, USA, 2004.

[4] D. Nace and M. Pióro, "Max-min fairness and its applications to routing and load-balancing in communication networks: A tutorial," *IEEE Communications Surveys and Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.

[5] I. Ahmed, A. Mohamed, and H. Alnuweiri, "On the fairness of resource allocation in wireless mesh networks: a survey," *Wireless Networks*, vol. 19, pp. 1451–1468, 2013.

[6] R. R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 18, no. 1, pp. 183–190, 1988.

[7] R. R. Yager and J. Kacprzyk, *The Ordered Weighted Averaging Operators, Theory and Applications*, Kluwer Academic, Dordrecht, The Netherlands, 1997.

[8] W. Ogryczak and T. Śliwiński, "On equitable approaches to resource allocation problems: the conditional minimax solutions," *Journal of Telecommunications and Information Technology*, vol. 3, pp. 40–48, 2002.

[9] M. M. Kostreva and W. Ogryczak, "Linear optimization with multiple equitable criteria," *RAIRO Operations Research*, vol. 33, no. 3, pp. 275–297, 1999.

[10] J. Rawls, *The Theory of Justice*, Harvard University Press, Cambridge, Mass, USA, 1971.

[11] D. E. Bell and H. Raiffa, "Risky choice revisited," in *Decision Making Descriptive, Normative and Prescriptive Interactions*, pp. 99–112, Cambridge University Press, Cambridge, UK, 1988.

[12] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*, Wiley, New York, NY, USA, 1994.

[13] W. Ogryczak, "Multiple criteria linear programming model for portfolio selection," *Annals of Operations Research*, vol. 97, pp. 143–162, 2000.

[14] W. Ogryczak and T. Śliwiński, "On solving linear programs with the ordered weighted averaging objective," *European Journal of Operational Research*, vol. 148, no. 1, pp. 80–91, 2003.

[15] V. Torra, "The weighted OWA operator," *International Journal of Intelligent Systems*, vol. 12, pp. 153–166, 1997.

[16] W. Ogryczak and T. Śliwiński, "On optimization of the importance weighted OWA aggregation of multiple criteria," in *Computational Science and Its Applications—ICCSA 2007*, vol. 4705 of *Lecture Notes in Computer Science*, pp. 804–817, Springer, Berlin, Germany, 2007.

[17] J. Hurkała and T. Śliwiński, "Fair flow optimization with advanced aggregation operators in Wireless Mesh Networks," in *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems*, pp. 415–421, 2012.

[18] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, pp. 23–30, 2005.

[19] M. Pióro, M. Zotkiewicz, B. Staehle, D. Staehle, and D. Yuan, "On max-min fair flow optimization in wireless mesh networks," *Ad Hoc Networks*, vol. 13, pp. 134–152, 2014.

[20] A. Capone, G. Carello, I. Filippini, S. Gualandi, and F. Malucelli, "Routing, scheduling and channel assignment in WMN networks: optimization, models and algorithms," *Ad Hoc Networks*, vol. 8, pp. 545–563, 2010.

[21] G. B. Dantzig and P. Wolfe, "The decomposition algorithm for linear programming," *Operations Research*, vol. 8, pp. 101–111, 1960.

[22] W. Ogryczak and T. Ślivińѕki, "On efficient WOWA optimization for decision support under risk," *International Journal of Approximate Reasoning*, vol. 50, no. 6, pp. 915–928, 2009.

[23] M. Grabish, S. A. Orlowski, and R. R. Yager, "Fuzzy aggregation of numerical preferences," in *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*, pp. 32–68, Kluwer Academic, Dodrecht, The Netherlands, 1999.

[24] S. J. Brams and A. D. Taylor, *Fair Division: From Cacke Cutting to Dispute Resolution*, Cambridge University Press, Cambridge, UK, 1996.

[25] W. Ogryczak, "On principles of fair resource allocation for importance weighted agents," in *Proceedings of the International Workshop on Social Informatics (SOCINFO '09)*, pp. 57–62, pol, June 2009.

[26] D. S. Lee, V. S. Vassiliadis, and J. M. Park, "A novel threshold accepting meta-heuristic for the job-shop scheduling problem," *Computers and Operations Research*, vol. 31, no. 13, pp. 2199–2213, 2004.

[27] D. S. Lee, V. S. Vassiliadis, and J. M. Park, "List-based threshold-accepting algorithm for zero-wait scheduling of multiproduct batch plants," *Industrial and Engineering Chemistry Research*, vol. 41, no. 25, pp. 6579–6588, 2002.

[28] C. D. Tarantilis and C. T. Kiranoudis, "A list-based threshold accepting method for job shop scheduling problems," *International Journal of Production Economics*, vol. 77, no. 2, pp. 159–171, 2002.

[29] C. D. Tarantilis, C. T. Kiranoudis, and V. S. Vassiliadis, "A list based threshold accepting algorithm for the capacitated vehicle routing problem," *International Journal of Computer Mathematics*, vol. 79, no. 5, pp. 537–553, 2002.

[30] C. D. Tarantilis, C. T. Kiranoudis, and V. S. Vassiliadis, "A list based threshold accepting metaheuristic for the heterogeneous fixed fleet vehicle routing problem," *Journal of the Operational Research Society*, vol. 54, no. 1, pp. 65–71, 2003.

[31] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, "Optimization by simulated annealing," *The American Association for the Advancement of Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[32] V. Černý, "Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, no. 1, pp. 41–51, 1985.

[33] C. Koulamas, "A survey of simulated annealing applications to operations research problems," *Omega*, vol. 22, no. 1, pp. 41–56, 1994.

[34] P. Tian, J. Ma, and D.-M. Zhang, "Application of the simulated annealing algorithm to the combinatorial optimization problem with permutation property: an investigation of generation mechanism," *European Journal of Operational Research*, vol. 118, no. 1, pp. 81–94, 1999.

[35] J. Hurkała and A. Hurkała, "Effective Design of the Simulated Annealing Algorithm for the Flowshop Problem with Minimum Makespan Criterion," in *Journal of Telecommunications and Information Technology*, vol. 2, pp. 92–98, 2012.

[36] J. Hurkała and A. Hurkała, "Fair optimization with advanced aggregation operators in a multicriteria facility layout problem," in *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems*, pp. 355–362, 2013.

[37] Y. M. Wang and C. Parkan, "A minimax disparity approach for obtaining OWA operator weights," *Information Sciences*, vol. 175, no. 1-2, pp. 20–29, 2005.

[38] B. S. Ahn, "Preference relation approach for obtaining OWA operators weights," *International Journal of Approximate Reasoning*, vol. 47, no. 2, pp. 166–178, 2008.

[39] G. R. Amin, "Notes on properties of the OWA weights determination model," *Computers and Industrial Engineering*, vol. 52, no. 4, pp. 533–538, 2007.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 625642, 9 pages
http://dx.doi.org/10.1155/2014/625642

Research Article

Max-Min Fairness in WMNs with Interference Cancelation Using Overheard Transmissions

Mateusz Żotkiewicz

Institute of Telecommunications, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warszawa, Poland

Correspondence should be addressed to Mateusz Żotkiewicz; mzotkiew@tele.pw.edu.pl

Received 28 February 2014; Accepted 24 April 2014; Published 13 May 2014

Academic Editor: Dritan Nace

We show an impact of using interference cancelation mechanisms for signals that have been overheard in the past on performance of fair wireless mesh networks. In our research we show that even in those very restricted conditions and max-min cost function, the idea of interference cancelation can significantly increase the capacity of such networks. In order to approximate possible advantages of using interference cancelation in the considered conditions, we propose a novel MIP model that allows for calculating perfect scheduling and maximal throughput in a network. We compare the results with cases when the interference cancelation mechanisms are disabled. Our results show that using interference cancelation mechanisms for signals that have been overheard in the past increases a network throughput by 40% on average in approximately 20% of test cases.

1. Introduction

Mesh networking in wireless communications is a way to increase network capacity, for example, with respect to single-hop cellular networks, through using short-distance radio links of relatively high capacity and establishing point-to-point connections on multi-hop routes. In this way packets from an origin are relayed by intermediate nodes on their way to a destination. Today, wireless mesh networks (WMN) are mainly used for providing affordable Internet access to communities of users in metropolitan areas (see [1, 2]). A typical wireless access network (WAN) is based on the IEEE 802.11-family Wi-Fi standards and consists of a set of fixed mesh nodes—routers and Internet gateways—interconnected by radio links. The mesh routers serve mesh clients (which are either fixed or mobile) and at the same time perform the packet relay function. WMN is a cost-efficient off-the-shelf networking that provides bandwidth in the range of up to hundreds Mbps. For popular surveys on WMN, see [3, 4].

Wireless PHY layer techniques are very mature nowadays and can achieve close to Shannon limit transmission capacities in point-to-point scenarios. Therefore, novel research directions that can result in the capacity increase of wireless networks are of interest now. One of such directions is to increase the spatial reuse by increasing a number of simultaneous transmissions that use the same medium. The direction proved to be very promising when IEEE 802.11n enhancements for higher throughput amendment [5] started to be widely used and a throughput of a single Wi-Fi transmission increased from 54 Mbps to 600 Mbps, mostly due to introduction of multiple-input multiple-output (MIMO) techniques.

In MIMO a source node transmits a number of independent data streams using a number of different antennas. In a sink node the streams are received also using a number of antennas. Although the streams are transmitted simultaneously and they strongly interfere with each other, they still can be successfully decoded due to phase shifts between them while being received by different antennas of the sink node.

The idea of using receiving techniques that involve recognizing and removing interferences from a signal is now shifting from a point-to-point scenario (like in MIMO) to more general scenarios like concurrent transmissions of many independent nodes. The technique is called *interference cancelation* and has been successfully implemented, for instance, in cellular networks [6]. A recent survey on successful interference cancelation implementations can be found in [7]. Currently, possibilities of employing the techniques in wireless mesh networks along with other ideas like zero-canceling [8] or interference alignment [9, 10] are being investigated.

One of assumptions of majority of work on interference canceling in wireless mesh networks is that canceling mechanisms are perfect. Therefore, if an interfering signal can be decoded, then it can be completely removed. The assumption works well when acceptable signal-to-interference-plus-noise (SINR) ratios of used modulations and codings schemes (MCSs) are rather low. When we assume that utilized MCSs require much higher SINR thresholds, then the quality of the interference canceling mechanisms becomes a factor and the previously justified assumption of perfect cancelation should be reconsidered.

In this paper we consider a case when SINR thresholds are relatively high and the usage of interference cancelation mechanisms is restricted only to cases when an interfering signal is canceled not because it can be decoded, but because it had been overheard in the past. In our research we show that even in this very restricted conditions the idea of interference cancelation can significantly increase the capacity of a wireless mesh network. In order to approximate possible advantages of using interference cancelation in the restricted conditions, we propose a novel mixed integer programming (MIP) model that describes the problem and allows for finding a perfect scheduling, thus calculating maximal throughput in a network. We compare the results with results obtained using an MIP model of [11] that describes a case when the interference cancelation mechanisms are not used. Our results show that using interference cancelation mechanisms in the restricted conditions increases a network throughput by 40% on average in approximately 20% of test cases.

The paper in organized as follows. In Section 2 we briefly explain technical issues of interference cancelation in overall and special features of the approach used in this research in detail. Next, in Section 3 we present a novel MIP model that allows for optimization of scheduling in wireless mesh networks that use the interference cancelation mechanisms for overheard transmissions. In Section 4, numerical results are presented. They are followed by conclusions given in Section 5.

2. Interference Canceling

The core components of a WMN that form its infrastructure/backbone are mesh points (MPs). They are responsible for relaying traffic. They can relay traffic of either other MPs or legacy nodes (STAs). Obviously they can also send their own traffic. Legacy STAs are connected to MAPs (specialized MPs that also act as access points for STAs) and follow the principle of a generic wireless local area network node. Other specialized MPs are responsible for connecting a WMN to an external network (usually the Internet). They are called mesh point portals or gateways and are abbreviated by MPP.

In radio networks two stations A and B can communicate if the power of the signal received from station A at station B in comparison to the noise at station B is greater than an acceptable *signal-to-noise* (SNR) threshold for the applied modulation and coding scheme (MCS). In WMN networks the radio resources are shared between all MPs. Thus, typically all MPs cannot transmit simultaneously but, of course, simultaneous transmissions by subsets of stations are still possible. In such a case, however, not only the noise can disrupt transmission but also the interferences induced by other MPs. Therefore, in the context of WMN we use the notion of *signal-to-interference-plus-noise ratio* (SINR) instead of SNR. SINR is understood as a proportion of a received power to a sum of the noise and the received interferences.

As there is little one can do with the noise, novel ways interferences that can be combated are of great interest. Two of such techniques, that is, zero-forcing [8] and interference alignment [9], are successfully used in point-to-point transmissions in MIMO technology. In WMN a promising interference combating strategy consists of various *interference cancelation* (IC) techniques. The notion behind IC is that if a node receives a signal consisting of an interfering signal A and a signal of interest B (assume that B cannot be decoded due to the interferences) and data encoded in A are known, then the interfering signal A can be reconstructed in the considered node using the known data, subtracted from the received signal, and finally B can be decoded.

The strategy can be used twofold. First, the signal A can be itself strong enough that it can be decoded in the receiving node regardless of interferences created by the signal B. Then, having data encoded in A, it is possible to reconstruct A and subtract it from the received signal. This approach was used in practice, for instance, in [12] where authors successfully implemented IC in a ZigBee network. However, the approach has one major disadvantage: it proved to work efficiently only with relatively low SINR thresholds. The reason is limited efficiency of IC detectors. In ideal conditions it is possible to reduce an interfering signal by up to 30 dB [12]. In practice, due to propagation and multipath issues, the efficiency rate is much smaller. Therefore, assuming that an SINR threshold is relatively high, if it is possible to decode an interfering signal A, then the signal is too strong to be canceled to such an extent that a signal of interest B can also be decoded. The way such a version of the problem can be approached in terms of optimization can be found in [13].

In our research we concentrate on the second way of using IC (see, for instance, [14, 15]). In this approach a main assumption is that data carried by an interfering signal A is known from previous transmissions. Consider an example presented in Figure 1, where a wants to send a packet to c and d wants to send a packet to f. Both packets are relied, the former by b, the latter by e. The arrows represent that possible links packets can be sent on; that is, if there is an arrow between a pair of MPs, then packets can be successfully sent between them. What is more, the arrows also represent possible interferences; for example, if a is transmitting to b, then d cannot transmit to e as the first transmission will jam the second. Therefore, in order to successfully meet both traffic demands, four time slots have to be used, because, due to possible interferences, there are no pairs of needed transmissions that can be conducted simultaneously.

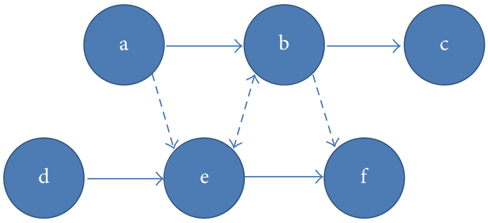


Figure 1: Example of using interference cancelation.

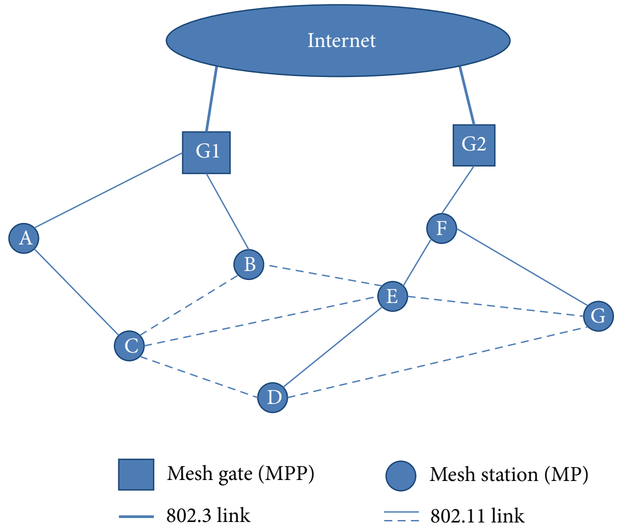


Figure 2: Wireless mesh network architecture.

However, when transmissions are properly ordered and IC mechanisms are utilized, the traffic demands can be met in just three time slots. First, a transmits to b, but the packet is overheard, that is, successfully received and decoded, by e. Then, b transmits to c and simultaneously d transmits to e. During this step e receives two signals (from b and from d) that interfere with each other. However, e knows data encoded in the signal from b; thus it can reconstruct it and subtract it from the received signal (perform interference cancelation). Finally, it can decode the packet sent from d. In the last step, e sends the decoded packet to f, thus satisfying the traffic demands. This simple example (other more sophisticated examples can be found in [16]) illustrates IC mechanisms that are considered in our research.

A vigilant eye would notice that this IC mechanism can be only used when a packet relayed by a node is not subject to any modifications in this node. Obviously that is usually not the case, as the relay node changes both a header and a CRC code that ends the packet. However, when appropriately approached, even this obstacle can be combated. The way of doing this is to let the to be canceled packet by transmitted slightly earlier. If the header of the packet is received by a node that is going to cancel it without any interferences, then the rest of this packet including CRC can be treated with IC mechanisms without any further problems and will not affect other transmissions that are supposed to happen right after the header of the canceled packet is received.

A major novelty of this paper is a tractable optimization model that encompasses the described IC mechanisms. In our research we assume a common interference model where two transmissions interfere only if a sender in the first transmission can successfully transmit to a receiver in the second transmission, or a sender in the second transmission can successfully transmit to a receiver in the first transmission. More sophisticated interference model, where a signal power is taken into account and all concurrent transmission interacts with each other, can be found, for instance, in [17].

We consider an architecture presented in Figure 2, where a set of MPs, a set of MPPs, and a set of links, routing, and possible interferences are given, and our task is to schedule transmissions in an optimal way. As volumes of demands in WMN are hard to predict, we decided to use the max-min fairness (MMF) paradigm (see, for instance, [18]) and maximize the throughput of the worst served MP.

We assume that the considered WMN is synchronized; a number of time slots are infinitely repeated. Our task is to assign point-to-point transmissions to those time slots in such a way that the objective is maximized. In order to better understand the notion of time slots, consider the example of Figure 1 with demands from a to c and from d to f and a bit rate of each link equal to 1 Mbps. If there are four available time slots and the IC mechanisms are disallowed, then both a and d will be the worst served MPs with 0.25 Mbps for each of them. What is more, when the IC mechanisms are used, the situation will not change. Although only three time slots are needed to serve each relevant MP with 0.25 Mbps, the fourth time slot cannot be effectively used; thus it is wasted in every cycle. On the other hand, when there are six available time slots and the IC mechanisms are disallowed, then only 0.17 Mbps can be provided to the worst served MP, while the other MP can be provided with 0.33 Mbps. However, when the IC mechanisms are used, both MPs can be served with 0.33 Mbps.

There is substantial literature on the optimization of scheduling and MCS selection in wireless mesh networks (see, for instance, [11, 17] and references therein). However, the topic of the scheduling and MCS selection optimization under IC constraints has not been yet exhaustively researched. In [19] the problem is raised, but the authors do not take routing into account. On the other hand, in [20] (extended later in [21, 22]), although routing is taken into account, the methodology of solving the problem is different and does not guaranty the optimality of solutions. What is more, in both cases perfect IC is assumed and overheard transmissions are not taken into account as far as IC is concerned.

3. Optimization Model

3.1. Notation. In order to formally present the problem, we introduce the following notation.

Indices

e: Mesh link

t: Time slot

m: MCS

v, w: Mesh nodes

d: Demand

(v, w): Mesh link from mesh node v to mesh node w

$a(e)$: Mesh node that is a source of mesh link e

$b(e)$: Mesh node that is a sink of mesh link e.

Sets

$\mathscr{V}$: Set of mesh nodes

$\mathscr{E}$: Set of mesh links

$\delta^+(e)$: Set of mesh links leaving mesh node e

$\delta^-(e)$: Set of mesh links entering mesh node e

$\mathscr{D}$: Set of demands

$\mathscr{P}_d$: Set of mesh links that form a path for demand d

$\mathscr{M}$: Set of MCSs

$\mathscr{T}$: Set of time slots.

Constants

B^m: Bit rate of MCS m

J^m_{ve}: 1 if mesh node v can jam a transmission at mesh link e that uses MCS m; 0 otherwise; notice that $J^m_{ve} = 0$ when $a(e) = v$

T^m_e: 1 if mesh link e can be successfully used with MCS m; 0 otherwise.

Variables

Y^m_{et}: 1 if mesh link e is transmitting in time slot t using MCS m; 0 otherwise

X_{vt}: 1 if mesh node v is transmitting in time slot t; 0 otherwise

y^m_{etv}: 1 if mesh link e is transmitting in time slot t using MCS m and the data is decoded in mesh node v; 0 otherwise; notice that the data can be decoded not only by $b(e)$

x^m_{vtm}: 1 if mesh node v is transmitting in time slot t using MCS m but mesh node w is canceling the interference; 0 otherwise

l_{vw}: Average transmission bit rate sent to mesh node v overheard by mesh node w

c_e: Average transmission bit rate at mesh link e

h_d: Average transmission bit rate for demand d

$h_{\min}$: Average transmission bit rate for the worst served demand.

In order to use the above notation, the defined sets and constants have to be relatively stable and possible to obtain. Therefore, the following requirements have to be met.

(i) $\mathscr{V}$ is known and does not change. When failures are not considered, the requirement is met.

(ii) $\mathscr{E}$ is known and does not change. When propagating conditions are relatively stable and nodes do not move, the requirement is met.

(iii) $\delta^+(e)$ is known and does not change. See above.

(iv) $\mathscr{D}$ is known and does not change. When a network consists of a number of MPPs and all other nodes are connected to them using static paths, the requirement is met.

(v) $\mathscr{P}_d$ serving a demand d (for all $d \in \mathscr{D}$) does not change. When a routing is static, the requirement is met.

(vi) $\mathscr{M}$ is known and does not change. The requirement is always met.

(vii) $\mathscr{T}$ is known and does not change. When a network is synchronized, the requirement is met. We assume that a number of possible time slots are given.

(viii) B^m are known and do not change. The requirement is always met.

(ix) J^m_{ve} are known and do not change. The requirement needs a detailed discussion; see below.

(x) T^m_e are known and do not change. In our research we compute constants T^m_e and J^m_{ve} using a mechanism described below. When properly used, the mechanisms guarantee that the requirements concerning T^m_e and J^m_{ve} are met.

An approach used to compute constants T^m_e and J^m_{ve} is similar to the one presented in [11]. In order to explain it in details the following notation is necessary.

N: Noise level

S_m: SINR threshold for MCS m

P_{vw}: Power received by mesh node w when mesh node v is transmitting.

The notation above does not appear in the final formulation of the problem and is solely used in order to define constants T^m_e and J^m_{ve}. The former constant is defined by (1), while the latter by (2). The equations guarantee that a given MCS can be used only if power of a signal of interest received at a node is greater than the noise and considered interferences. It is worth to mention that constants S_m are in linear scale here and not logarithmic:

$$T^m_e = \begin{cases} 1 & \text{if } \dfrac{P_{a(e)b(e)}}{N} \geq S_m \\ 0 & \text{if } \dfrac{P_{a(e)b(e)}}{N} < S_m, \end{cases} \tag{1}$$

$$J^m_{ve} = \begin{cases} 1 & \text{if } \dfrac{P_{a(e)b(e)}}{N + P_{vb(e)}} < S_m \\ 0 & \text{if } \dfrac{P_{a(e)b(e)}}{N + P_{vb(e)}} \geq S_m. \end{cases} \tag{2}$$

If we assume that mesh nodes do not move and propagating conditions are relatively stable, then N and P_{vw} can be treated as constants in both (1) and (2). As S_m is also a constant, therefore T_e^m and J_{ve}^m can also be treated as constants.

Finally, we assume that a kind of OSPF routing is used. The assumption is irrelevant from the point of view of the notation. However, in our model we require that a path used by a packet before reaching a given node is irrelevant. The requirement is met when there is exactly one path to each node; thus it can be met when a kind of OSPF routing is used in a network.

To summarize, in order to use the model presented in this paper, the following conditions have to be assumed.

(i) Failures are not considered.

(ii) Mesh nodes do not move.

(iii) Propagating conditions are relatively stable.

(iv) Network consists of a number of MPPs and all other nodes are connected to them using static paths.

(v) Routing is static.

(vi) Network is synchronized.

(vii) Kind of OSPF routing is used.

3.2. MIP Model. When the conditions presented above are assumed, we can formally state the problem in the form of formulations (3a)–(3m) below. It is a compact mixed integer-linear programming formulation that has proved to be tractable for network consisting of tens of nodes:

maximize $h_{\min}$
subject to

$$\sum_{m\in\mathscr{M}}\sum_{e\in\delta^+(v)} Y_{et}^m = X_{vt} \quad \forall v\in\mathscr{V},\ \forall t\in\mathscr{T}, \tag{3a}$$

$$y_{etv}^m \le Y_{et}^m \quad \forall e\in\mathscr{E},\ \forall t\in\mathscr{T},\ \forall v\in\mathscr{V},\ \forall m\in\mathscr{M}, \tag{3b}$$

$$l_{vw} = \frac{1}{|\mathscr{T}|}\sum_{m\in\mathscr{M}} B^m \sum_{t\in\mathscr{T}}\sum_{e:b(e)=v} y_{etw}^m \quad \forall v,w\in\mathscr{V}, \tag{3c}$$

$$l_{vw} \ge \frac{1}{|\mathscr{T}|}\sum_{m\in\mathscr{M}} B^m \sum_{t\in\mathscr{T}} x_{vtw}^m \quad \forall v,w\in\mathscr{V}, \tag{3d}$$

$$x_{vtw}^m \le \sum_{e\in\delta^+(v)} Y_{et}^m \quad \forall v,w\in\mathscr{V},\ \forall m\in\mathscr{M},\ \forall t\in\mathscr{T}, \tag{3e}$$

$$y_{etw}^m + X_{vt} - \sum_{m'\in\mathscr{M}} x_{vtw}^{m'} \le 2 - J_{v(a(e),w)}^m \quad \forall e\in\mathscr{E}, \quad \forall t\in\mathscr{T},\quad \forall m\in\mathscr{M},\quad \forall v,w\in\mathscr{V}, \tag{3f}$$

$$y_{etv}^m \le T_{(a(e),v)}^m \quad \forall e\in\mathscr{E},\ \forall t\in\mathscr{T},\ \forall m\in\mathscr{M},\ \forall v\in\mathscr{V}, \tag{3g}$$

$$c_e = \frac{1}{|\mathscr{T}|}\sum_{m\in\mathscr{M}} B^m \sum_{t\in\mathscr{T}}\left(y_{etb(e)}^m - x_{a(e)tb(e)}^m\right) \quad \forall e\in\mathscr{E}, \tag{3h}$$

$$x_{vtv}^m = 0 \quad \forall v\in\mathscr{V},\ \forall t\in\mathscr{T},\ \forall m\in\mathscr{M}, \tag{3i}$$

$$\sum_{d:e\in\mathscr{P}_d} h_d \le c_e \quad \forall e\in\mathscr{E}, \tag{3j}$$

$$h_{\min} \le h_d \quad \forall d\in\mathscr{D}, \tag{3k}$$

$$X_{vt}, Y_{et}^m, y_{etv}^m, x_{vtw}^m \in \{0,1\}, \tag{3l}$$

$$l_{vw}, c_e, h_d, h_{\min} \ge 0. \tag{3m}$$

In our research we use the max-min cost function [23]; thus the objective of (3a)–(3m) is to maximize the minimum flow $h_{\min}$ to the worst served MP. Constraint (3a) assures that a node during one time slot can transmit only on one link using only one MCS. Constraint (3b) prevents nodes from decoding data that are not transmitted. In other words, only packets that have been sent can be decoded. The purpose of (3c) is to calculate the average transmission bit rate sent to $b(e)$ (w in the next constraint) and decoded by v. The computed value is an upper bound for the average transmission bit rate sent from w and canceled in v. The upper bound is enforced by (3d). Obviously the interference cancelation can take place only if a transmission to be canceled takes place. This condition is enforced by (3e). The next constraint, that is, (3f), assures that a transmission can be successfully decoded only if it is not jammed by other transmissions. Notice that transmissions that have been earlier overheard and currently are canceled cannot jam concurrent transmissions. Constraint (3g) assures that only MCSs satisfying SNR thresholds are used, while (3h) calculates the average useful transmission bit rate at each link. Notice that canceled transmissions do not count towards the average useful transmission bit rate. Constraint (3i) disallows MPs to cancel self-interferences. It is worth noticing that some work on canceling self-interferences exists and a full-duplex has even been implemented in practice [24]. However, it required a special technique called *antenna cancelation*; thus in our research we assume that the self-interferences cancelation is impracticable and is not being used. Constraint (3j) divides available capacity of a network between different traffic demands, and (3k) assures that the division is fair; average throughput of the worst served demand is maximized. Finally, constraints (3l) and (3m) are responsible for binarity and nonnegativity of the variables.

3.3. Valid Inequalities. The presented MIP model is relatively complicated and solving it requires significant computation effort. In order to increase its applicability, we present three practical valid inequalities. They are not indispensable from the point of view of the correctness of the model. However, they tight a polytop of a relaxed problem, thus making branch-and-bound approach work more efficiently:

$$\sum_{t\in\mathscr{T}}\sum_{m\in\mathscr{M}}\left(y_{etb(e)}^m - x_{a(e)tb(e)}^m\right) \ge 1 \quad \forall e\in\mathscr{P}_d,\ \forall d\in\mathscr{D}, \tag{4a}$$

Table 1: Modulation and coding schemes.

Modulation	Coding	Throughput	SINR [log.]	SINR [lin.]
BPSK	1/2	6.0 Mbps	3.48 dB	2.23
QPSK	1/2	12.0 Mbps	6.63 dB	4.6
16-QAM	1/2	24.0 Mbps	12.76 dB	18.88
64-QAM	1/2	48.0 Mbps	20.29 dB	106.9

$$X_{vt} + \sum_{e\in\mathscr{E}} \sum_{m\in\mathscr{M}} \left(y^m_{etv} - x^m_{a(e)tv} \right) \leq 1 \quad \forall v \in \mathscr{V},\ \forall t \in \mathscr{T}, \tag{4b}$$

$$\sum_{e\in\delta^-(v)} \sum_{t'\in\mathscr{T}} \sum_{m'\in\mathscr{M}} y^{m'}_{et'v} \geq x^m_{vtw} \quad \forall v, w \in \mathscr{V},\ \forall t \in \mathscr{T},\ \forall m \in \mathscr{M}. \tag{4c}$$

The first valid inequality (4a) guarantees that on each mesh link that is used by at least one demand there is a transmission scheduled during at least one time slot. What is more, during that time slot not only the transmission has to occur but also it cannot be subject to interference cancelation at the receiving end of the mesh link. The inequality limits the number of possibilities of dividing time slots into smaller parts while solving relaxed versions of the problem during the B&B process. Notice that this inequality is formally not a *valid* inequality, as it cuts feasible solutions of cost 0. However, from the practical point of view, those solutions are useless. Therefore, we decided to treat this inequality as valid.

The second valid inequality (4b) guarantees that a mesh node either transmits or receives data from aother mesh node. The inequality disallows a number of relaxed solutions when a mesh node receives and decodes a number of "half" transmissions and simultaneously "half" transmits itself; that is, $y^m_{etv} = 0.5$ and $X_{vt} = 0.5$ in constraint (3f).

The last valid inequality (4c) guarantees that if a mesh node performs interference cancelation during at least one time slot, then it had to overheard a transmission to be canceled during at least one other time slot. The inequality introduces tight connections between binary variables $y^{m'}_{et'v}$ and x^m_{vtw}. In model (3a)–(3m) the variables were related using variable l_{vw} and constraints (3c) and (3d).

4. Numerical Results

We tested numerically the optimization model presented in Section 3 using CPLEX 12.5 set up on Intel Core 2 Quad CPU Q6600 2.40 GHz with 4.00 GB RAM. We used 50 randomly generated network topologies for each test case from Table 2. All nodes of the networks were randomly located on a square. We followed [25] and assumed the path loss exponent of 4 while calculating power received at each station and establishing mesh links. Notice that the topology and the path loss exponent allow us to calculate both T^m_e and J^m_{ve}. Each MP is routed to the nearest MPP using the shortest path with respect to weights of links. We set the weight of link e to the inverse of $P_{a(e)b(e)}$. Each link uses an available (satisfying the SINR constraint) MCS from Table 1 with maximal B^m. In order to minimize a number of infeasible instances, locations of nodes were scaled in such a way that the longest utilized link in a network was of a maximal length allowing for a successful transmission using the least demanding MCS.

Table 2: Considered network topologies.

Topology	MPPs	MPs	Avg. mesh links
N6	5	1	15.0
N8	7	1	25.0
N10	8	2	32.5
N12	10	2	42.3
N14	11	3	54.0
N16	13	3	61.4
N18	15	3	71.4
N20	16	4	83.5
N22	18	4	97.5
N24	19	5	104.2
N26	21	5	114.1
N28	23	5	131.8
N30	24	6	143.4

Numerical results are presented in Table 3. The first column contains topology identifiers. They inform which test case from Table 2 is under consideration. The second column contains the number of time slots in a test case. The following six columns contain efficiency specific information. For each case we gathered mean running time of the MIP solver, a number of timeouts (the time limit was set to ten minutes), and a mean number of B&B nodes entered, for both a case with valid inequalities of Section 3.3 and a case without the valid inequalities. The last two columns justify usage of IC in WMN. The first of them contains a percentage of cases which can be improved by employing IC, while the second contains an average obtained gain for those cases. It is expressed in percents and understood as $(B - A)/A$ times 100%, where A is $h_{\min}$ for a test case that does not use IC and B is $h_{\min}$ for a test case taking advantage of IC.

The numerical results prove the expected complexity of the problem. Running times show a limited tractability of the presented MIP model. However, the limited tractability is accompanied by considerable gains in terms of network capacity. It is worth mentioning that the max-min cost function used in our research is one of the most restrictive and in many cases it is impossible to obtain better results due to one very inconveniently located mesh node. In those cases, although the max-min gain is equal to zero, using IC techniques still can increase throughput for better served mesh nodes.

As for the usage of the valid inequalities of Section 3.3, it almost always decreases a number of visited B&B nodes. However, it not always assures that a problem will be solved faster. In fact, as can be seen in Figure 3, it is usually profitable to use the valid inequalities for cases with a small number of time slots. However, when the number of time slots increases, then the valid inequalities lose their importance.

Running times presented in Figure 3 may create an impression that the model is useless for larger networks. Additionally taking into account the information about timeouts presented in Table 3, one may claim that with so

TABLE 3: Numerical results.

Top.	$\|\mathcal{T}\|$	Without valid inequalities			With valid inequalities			Improved [%]	Gain [%]
		Time [s]	#EOT	#B&B	Time [s]	#EOT	#B&B		
N6	8	0.22	0	7.48	0.3	0	0.6	4.0	46.7
N6	12	0.3	0	3.42	0.42	0	43.2	6.0	23.3
N8	8	0.56	0	36.54	0.76	0	18.2	18.0	35.6
N8	12	0.74	0	91.64	1.32	0	15.2	12.0	28.9
N10	8	0.86	0	60.16	1.66	0	32.0	14.0	53.8
N10	12	13.24	1	693.38	4.46	0	131.1	16.0	45.5
N12	8	1.64	0	143.02	3.4	0	42.1	20.0	56.1
N12	12	19.64	1	1093.7	22.76	1	943.4	24.0	28.3
N14	8	1.9	0	193.82	2.36	0	45.5	16.0	73.1
N14	12	2.98	0	199.96	7.3	0	242.1	12.0	22.1
N16	8	15.68	1	616.06	6.92	0	73.3	12.0	45.7
N16	12	16.26	0	471.9	29.22	0	532.2	28.0	25.9
N18	8	17.84	1	3065.92	8.34	0	209.1	20.0	45.5
N18	12	33.22	2	2278.5	28.4	1	517.0	36.0	25.3
N20	8	54.88	4	3142.44	31.14	1	280.3	8.0	22.0
N20	12	78.56	5	2479.86	76.32	4	990.6	22.0	30.9
N22	8	36.04	2	4182.44	20.26	0	334.3	22.0	47.1
N22	12	98.4	5	7991.82	95.18	4	1196.5	28.0	29.2
N24	8	58.22	4	5092	18.36	0	308.6	20.0	71.9
N24	12	81.1	5	3892.46	84.62	5	1651.6	24.0	30.5
N26	8	63.36	4	11441.86	26.4	0	581.9	18.0	53.9
N26	12	110.48	6	5597.08	127.74	6	2314.8	32.0	25.8
N28	8	64.5	3	4092	46	0	802.3	16.0	41.1
N28	12	188.84	12	7933.38	207.88	13	3962.1	34.0	32.3
N30	8	157.5	10	16574.44	55.44	1	868.8	20.0	54.7
N30	12	192.08	9	13423.08	197.5	11	3167.3	26.0	28.1
Average		50.35	2.88	3646.1	42.48	1.81	742.5	19.54	39.35

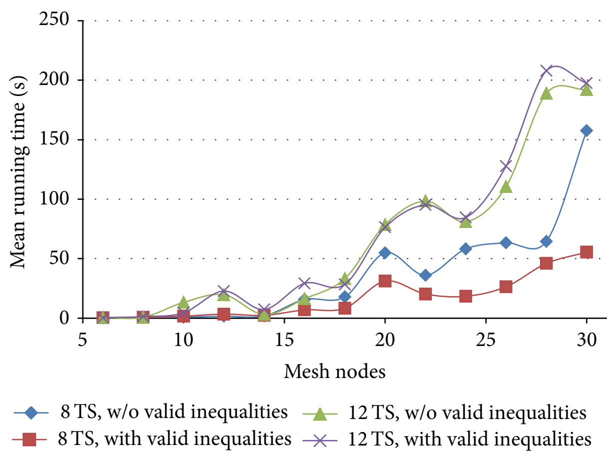


FIGURE 3: Running times.

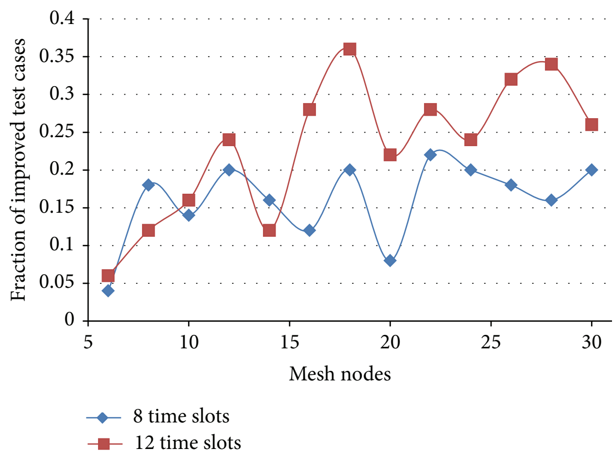


FIGURE 4: Fraction of improved test cases.

many unsolved test cases the model is impractical. However, notice that a timeout does not mean that the MIP solver has not returned any solution for a test case. A timeout only means that the MIP solver was not able to prove an optimality of a returned solution, which still in many cases was better than a solution obtained when IC was not considered.

Finally, our results show an interesting impact of increasing the number of time slots on the problem. As expected, increasing a number of time slots complicates the problem

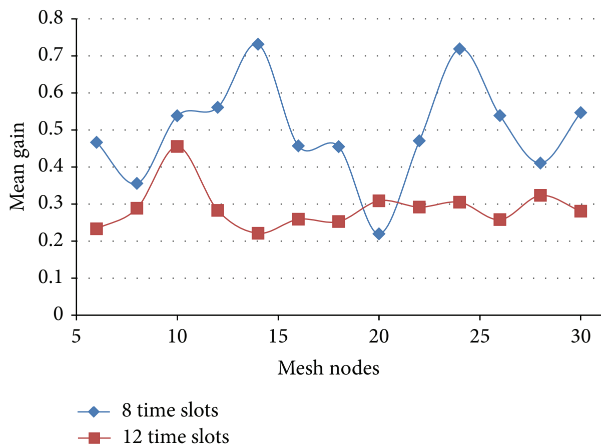


Figure 5: Mean gain.

making it more difficult to solve. This fact is clearly seen in Figure 3. However, an impact of increasing the number of time slots on efficiency of IC is not so straightforward. On one hand, as seen in Figure 4, the more the time slots are available, the bigger the chance that using IC can be profitable. On the other hand, as seen in Figure 5, the more the time slots are available, the smaller the gain IC can bring. Such a behavior is a result of an increased elasticity brought by an increased number of available time slots. The increased elasticity extends a feasibility region of our problem making it easier to find better solutions than those available in a case when IC is disallowed. On the other hand, with increased elasticity, all the changes in a solution result in much smaller impact on the global objective function, thus making possible gains smaller.

5. Conclusion

In the paper we show that the idea of interference cancelation can significantly increase the capacity of a wireless mesh network even when the max-min objective function is taken into account, and the interference cancelation can be applied only to signals that have been overheard in the past. In order to approximate possible advantages of using interference cancelation in the considered conditions, we presented a novel MIP model that allows for calculating perfect scheduling and maximal throughput in a network. We compare our results with results obtained for cases when the interference cancelation mechanisms are disabled. Our results show that using interference cancelation mechanisms for signals that have been overheard in the past increases a network throughput by 40% on average in approximately 20% of test cases.

The work presented in this paper can be extended in many different directions. One of possible extensions is to consider the full interference model instead of the used simplified interference model. Another interesting research direction is to present the problem using a noncompact formulation based on the idea of compatible sets (see [11, 17]) that allows for an indirect consideration of an infinite number of time slots. Notice that both of the suggested possible extensions cannot be directly incorporated into our MIP model and require deeper studies.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The work presented in this paper has been funded by National Science Center (Poland) under Grant 2011/01/B/ST7/02967 "integer programming models for joint optimization of link capacity assignment, transmission scheduling, and routing in fair multicommodity flow networks."

References

[1] "Bay area wireless users group," http://www.bawug.com/.

[2] "Seattle wireless," http://www.seattlewireless.net/.

[3] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, no. 9, pp. S23–S30, 2005.

[4] R. Bruno, M. Conti, and E. Gregori, "Mesh networks: commodity multihop ad hoc networks," *IEEE Communications Magazine*, vol. 43, no. 3, pp. 123–131, 2005.

[5] IEEE Standard for Information Technology, "Telecommunications and information exchange between systems-local and metropolitan area networks-specific requirements. Part 11: wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications amendment 5: enhancements for higher throughput," IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 As Amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009), 2009.

[6] J. G. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *IEEE Wireless Communications*, vol. 12, no. 2, pp. 19–29, 2005.

[7] N. I. Miridakis and D. D. Vergados, "A survey on the successive interference cancellation performance for single-antenna and multiple-antenna OFDM systems," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 312–335, 2013.

[8] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, New York, NY, USA, 2005.

[9] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, 2008.

[10] S. Gollakota, S. D. Perli, and D. Katabi, "Interference alignment and cancellation," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 159–170, 2009.

[11] M. Pióro, M. Zotkiewicz, B. Staehle, D. Staehle, and D. Yuan, "On max-min fair flow optimization in wireless mesh networks," *Ad Hoc Networks*, vol. 13, part A, pp. 134–152, 2014.

[12] D. Halperin, T. Anderson, and D. Wetherall, "Taking the sting out of carrier sense: interference cancellation for wireless LANs," in *Proceedings of the 14th ACM Annual International Conference on Mobile Computing and Networking (MobiCom '08)*, pp. 339–350, San Francisco, Calif, USA, September 2008.

[13] D. Yuan, V. Angelakis, L. Chen, E. Karipidis, and E. G. Larsson, "On optimal link activation with interference cancelation in wireless networking," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 2, pp. 939–945, 2013.

[14] B. Bandemer, Q. Li, X. E. Lin, and A. Paulraj, "Overhearing-based interference cancellation for relay networks," in *Proceedings of the 70th IEEE Vehicular Technology Conference Fall (VTC-Fall '09)*, pp. 1–5, Anchorage, Alaska, USA, September 2009.

[15] S. Zhang, S.-C. Liew, and H. Wang, "Blind known interference cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 8, pp. 1572–1582, 2013.

[16] L. E. Li, R. Alimi, D. Shen, H. Viswanathan, and Y. R. Yang, "A general algorithm for interference alignment and cancellation in wireless networks," in *Proceedings of the 29th IEEE International Conference on Computer Communications (INFOCOM '10)*, pp. 1–9, San Diego, Calif, USA, March 2010.

[17] A. Capone, G. Carello, I. Filippini, S. Gualandi, and F. Malucelli, "Routing, scheduling and channel assignment in Wireless Mesh Networks: optimization models and algorithms," *Ad Hoc Networks*, vol. 8, no. 6, pp. 545–563, 2010.

[18] M. Pióro and D. Medhi, *Routing, Flow, and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, 2004.

[19] V. Angelakis, L. Chen, and D. Yuan, "Optimal and collaborative rate selection for interference cancellation in wireless networks," *IEEE Communications Letters*, vol. 15, no. 8, pp. 819–821, 2011.

[20] S. Lv, W. Zhuang, X. Wang, and X. Zhou, "Scheduling in wireless ad hoc networks with successive interference cancellation," in *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pp. 1287–1295, Shanghai, China, April 2011.

[21] S. Lv, W. Zhuang, X. Wang, C. Liu, and X. Zhou, "Maximizing capacity in the SINR model in wireless networks with successive interference cancellation," in *Proceedings of the IEEE International Conference on Communications (ICC '11)*, pp. 1–6, Kyoto, Japan, June 2011.

[22] S. Lv, W. Zhuang, X. Wang, and X. Zhou, "Link scheduling in wireless networks with successive interference cancellation," *Computer Networks*, vol. 55, no. 13, pp. 2929–2941, 2011.

[23] W. Ogryczak, M. Pióro, and A. Tomaszewski, "Telecommunications network design and max-min optimization problem," *Journal of Telecommunications and Information Technology*, vol. 3, pp. 43–56, 2005.

[24] J. Choi II, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proceedings of the 16th ACM Annual Conference on Mobile Computing and Networking (MobiCom '10)*, pp. 1–12, September 2010.

[25] A. Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 948521, 11 pages
http://dx.doi.org/10.1155/2014/948521

Research Article

Constructing Fair Destination-Oriented Directed Acyclic Graphs for Multipath Routing

Katarzyna Kalinowska-Górska and Fernando Solano Donado

Faculty of Electronics and Information Technology, Warsaw University of Technology, ul. Nowowiejska 15/19, 00-665 Warsaw, Poland

Correspondence should be addressed to Katarzyna Kalinowska-Górska; k.kalinowska@stud.elka.pw.edu.pl

Received 7 February 2014; Accepted 24 March 2014; Published 28 April 2014

Academic Editor: Hanan Luss

Extensive research in the field of telecommunications has been done on the techniques of multipath routing, as they offer many advantages over conventional single-path routing methods. Some of these techniques make use of the so-called Destination-Oriented Directed Acyclic Graphs (DODAGs) which are constructed on the networks, usually in a distributed way. However, while defining methods of forming DODAGs, the authors of multipath algorithms tend to overlook a possibly significant issue which could, in a way, define the quality of a given DODAG in the context of multipath routing, namely, providing an equitable distribution of the paths between the nodes in the newly created DODAG. In this paper, a few requirements for constructing a "fair" DODAG are identified in the context of multipath routing. An optimization algorithm that tries to find an equitable solution according to these requirements is also presented. Three DODAG-creation algorithms that appear in the literature are simulated and compared against this equitable solution, and none of them is getting close to it in terms of fairness in the distribution of the paths. Moreover, two interesting properties of equitable solutions are revealed in the simulations.

1. Introduction

A Destination-Oriented Directed Acyclic Graph (DODAG) is a term used in [1] to describe a directed acyclic graph with exactly one root, where a root is a node which has no outgoing edges. Diverse multipath routing algorithms make use of DODAGs, such as [2–8]. While conventional single-path routing techniques form directed spanning trees rooted at the destination, multipath algorithms use DODAGs to provide one or more paths between a node-root pair (A directed tree is also a DODAG, but of a more restricted form.). The networks and DODAGs considered in this paper are simple graphs. As an example, a simple DODAG is shown in Figure 1.

It is worth noting that in a full multipath routing procedure two different tasks have to be performed: (1) determining a set of paths per node-destination pair and (2) distributing the flow on these paths. In some algorithms these tasks are easily separable, like in the case of equal-cost multi-path (ECMP), where the demand from the source to the destination is split uniformly between all equal-cost shortest paths. In other algorithms, such as iterative gradient minimization algorithms [9], these two steps merge and cannot be executed separately.

In scenarios where network topology changes are infrequent and the amount of energy is limited, for example, in wireless sensor networks for environmental monitoring, separating these two tasks could yield better overall performance, considering the energy consumed for calculations and transmission. By neglecting small changes in the topology, CPU cycles required for recalculating the set of paths (first task) can be spared and the number of signaling messages exchanged between the nodes can be reduced. In addition, total node's memory requirements can be reduced, as only one task needs to be tackled at a time. Therefore, smaller chipsets (with less energy consumption per cycle) can be employed.

Still, which of the two strategies, that is, executing both tasks jointly or separately, performs better, depends much on the particular application and is left out of the scope of this paper. Here, we consider such applications where both tasks can be solved separately without a substantial

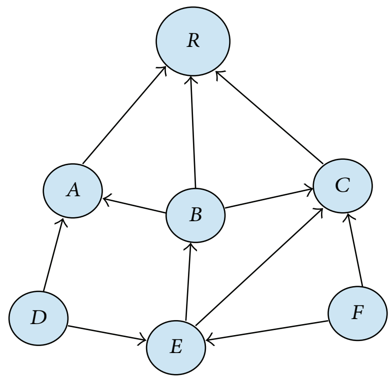


Figure 1: An example DODAG with the root at node R.

decrease in performance. In particular, we assume that a set of paths is fixed for long periods of time and flow reallocation procedures are executed more often, for example, once link qualities, nodes' capabilities or congestion state in the network change. As a considerable amount of literature has already been published on flow allocation algorithms, in this paper, we focus on the first task: how to find a large set of paths that can be used for long periods of time and which provides enough flexibility, taking into account possible flow reallocations. This can be achieved by maintaining a once-constructed DODAG in the network, with occasional local and global repairs when needed, as in [1].

Many different DODAGs might be constructed in the network for multipath routing. Hence, an obvious question arises; that is, what properties should the constructed DODAG have? The objectives of multipath routing include, among other things, providing resilience against node/link failures, minimizing the risk of link overload, or balancing the energy consumption of battery-powered nodes in case of Low-Power and Lossy Networks (LLNs). On the basis of these (and other) objectives, a conclusion can be made that during the lifetime of a single DODAG, a node would probably need a few paths for path switching, flow splitting, or introducing flow redundancy. This means that, when considering two nodes with a similar distance to the destination, it is better when both have several paths to the root in the DODAG than one having tens of paths and the other with only one or two paths available. This observation leads to an objective for the DODAG construction algorithm, which is to provide a *fair* distribution of paths per node-root pair in the DODAG. Obviously, this objective is strongly constrained by the network topology; in particular, it is never possible to obtain equal, other than one, number of paths in every node. This is due to the fact that in a simple DODAG there always exists at least one node which has exactly one path to the root. This is stated as Proposition 1 in the next section and proved in Appendix B.

There are more reasons why the *fairness* objective mentioned above does not simply mean trying to equalize the numbers of paths per node in the resulting DODAG. First of all, leveling down is undesired, which means that apart from equalizing the numbers of paths, their simultaneous maximization is required. Second, in most cases, the average number of paths of a node in the DODAG depends on its distance from the root. Typically, it grows as the distance increases (unless many long roundabout paths exist in the DODAG that lead from nodes closer to the root through nodes farther from the root). Third, the quality of considered paths is important. In particular, it could be of great advantage to reduce the length/cost of the paths used for routing and/or to provide node/link disjointness of the paths in the path sets [10]. Moreover, the number of paths in a graph grows exponentially as a function of the number of links. Merely enumerating all the paths would pose a serious computational problem, even for relatively small networks (e.g., of less than 50 nodes), which means that it is not feasible to consider all of them in reasonable time.

These observations imply that a solution is needed as follows:

(i) it tries to equalize, without leveling down, the numbers of paths per node at the same minimum hop count distance from the root (at the same "level");

(ii) when trying to provide an equitable distribution of the paths, it does not take all of them into consideration but reasonably selects subsets of all possible paths instead based on their properties, for example, their length.

Which properties of the paths selected for equalization are the most important depends on the particular application. Therefore, a solution is proposed in this paper which does not define *a priori* the desired path sets per node but allows potential users to define their own candidate path sets, one set per node, which are appropriate in a particular situation. An equitable distribution of the paths per node would then be provided, only taking into account the given path sets. These path sets might satisfy some predefined requirements; for example, they might represent shortest paths sets or disjoint paths sets or include other paths which are, for some reason, of interest to the user.

It is proved in Appendix A that even a simple problem of maximizing the number of paths from a single node to the DODAG root is NP-complete. Therefore, to achieve an equitable and yet efficient distribution of given paths, an IP optimization problem with a lexicographic maximin objective has been formulated, and an appropriate algorithm from [11] for solving this problem has been chosen. Solving the optimization problem results in a simple DODAG rooted at the destination node, with equitable distribution of input paths among all the nodes.

Of course, a constructed DODAG will typically contain more paths than just the ones belonging to the set provided by the user for optimization. These additional paths might be used for flow allocation as well if needed; for example, they might serve as backup paths when all of the input paths defined for a node are affected by a link/node failure. However, it is assumed that since they were not included by the user in the input path sets, they might be of "worse quality" (e.g., greater cost). In any case, since the number of additional paths of a node is not controlled by the optimization algorithm, their distribution will probably be random and no *fairness* will be achieved in general. On the other hand, for high enough levels, so many paths will be available per node that trying to obtain a general fair distribution makes little sense, since each node will get more paths than it will need anyway. What is important is providing each node with the best paths, as far as possible, which, again, justifies the approach taken in this paper. As for the nodes at lowest levels, if the input path sets are reasonably chosen, any additional paths in the DODAG will be very roundabout, therefore not very useful.

Authors of some of the DODAG-creation algorithms appearing in the literature, like LMR and TORA [5, 6], ignore the number and properties of paths available in the resulting DODAGs. In the IDAGs approach [8], two node- or link-independent DAGs are constructed, guaranteeing each node to have at least two node- or link-disjoint paths. MARA-MC [7] is proved by the authors to compute a large number of paths for a large fraction of source-destination pairs, when executing a DODAG construction procedure for every destination node. Authors of the MDVA algorithm [2] only check its performance (i.e., message overhead and convergence time) and do not consider properties like the number of paths at all. However, MDVA and other LFI-based algorithms by the same authors [3, 4] are compared to MARA-MC in [7] and proved to yield worse results regarding the average number of paths in the constructed DODAGs.

The remaining part of the paper is organized as follows. Section 2 describes the formulation of the optimization problem and the algorithm chosen for its solution. Section 3 presents the simulations and a comparison of three algorithms from the literature with the solution obtained by solving the lexicographic problem when providing sets of k shortest paths per node. Moreover, additional properties of the most equitable solutions are shown. Section 4 concludes the paper. Proofs for statements included in the paper are presented in Appendices A and B. Appendix C contains parameter values of the algorithm used for the generation of test networks.

2. Optimization Problem with Lexicographic Maximin Objective: Formulation and Algorithm

The problem of equitable distribution of the paths among the nodes in the graph corresponds in fact to a resource allocation problem [11], where different activities compete over limited resources which have to be allocated *fairly* among them. In this case, nodes correspond to activities and a single objective function of a node expresses the number of paths "allocated" to that node. The resources which are limited, or rather constrained by the network topology, are the directions of the links. The way of assigning these directions determines the set of paths available in the DODAG.

A lexicographic maximin objective [11] has been chosen for this problem. It can be viewed as an extension of a simple maximin objective. In this approach, first the number of paths for the node with fewest paths is maximized. Then, the number of paths for the node with second fewest paths is maximized, but without decreasing the smallest value calculated in the first step. Next, the number of paths for the node with third fewest paths is maximized, without decreasing the value of the first two and so forth. Hence, solving a lexicographic maximin problem means finding a solution vector (of numbers of paths of the nodes) which is lexicographically largest of all feasible vectors, when the values in the vectors are sorted in a nondecreasing order. It is not known in advance which node will occupy which lexicographic position in the solution vector. A lexicographic maximin (or minimax) solution is Pareto optimal.

It is worth mentioning here that a simple maximin objective would achieve nothing due to the following fact.

Proposition 1. *In a simple DODAG, there will always exist at least one node with exactly one path to the root.*

It will be one of the root's neighbours. Therefore, the minimum number of paths of all nodes in the DODAG will always be equal to one. If only predefined sets of paths are considered, the minimum number of input paths of all nodes in the DODAG will never be more than one. The proof of Proposition 1 is presented in Appendix B.

Applying lexicographic maximization to the DODAG construction problem with input path sets ensures that at least one candidate path per node survives in the DODAG, as long as the following two conditions are met:

(i) at least one path per node is inserted into the formulation;

(ii) in the provided path sets, there is at least one combination of paths containing at least one path per node where no paths exclude each other due to conflicting edge directions; that is, all these paths can coexist in a DODAG.

At least one path per node will be included in the DODAG due to the fact that lexicographic maximization will not allow a zero value in the first position in the solution vector, if it is possible to obtain a minimum equal to 1. The above conditions are met, for example, when providing nodes' shortest paths to the root (as they form a tree).

2.1. The Formulation. The generic network inserted into the formulation (No specific type of network is of interest, although the simulations were performed on random *ad hoc* networks.) is modeled as a simple, strongly connected directed graph $G = (N, E)$, where N is the set of nodes and E is the set of links. In addition, if there exists an edge from vertex m to vertex n, there also exists an edge in the reverse direction (from vertex n to vertex m); that is, the graph is symmetric. A set of candidate paths $p = 1, 2, \dots, P$ is provided, each path represented a sequence of C_p directed edges. The identifier of the destination (root) node is denoted by r. Neighbours of node m or r are indexed by n. The following variables are present in the formulation:

(i) $y_{mn} = 1$ if the (m, n) edge is included in the solution (output DODAG), 0 otherwise;

(ii) $Y_p = 1$ if a candidate path p is included in the DODAG, 0 otherwise;

(iii) U_m is an integer variable representing the number of paths from a node m to the root in the DODAG but only counting the paths included in the P set inserted into the formulation. No other paths available in the DODAG are counted by this variable.

The problem is formulated as follows:

$$\operatorname{lexmax} \left\{U^{(m)}(Y) = \left[U_{m1}(Y), U_{m2}(Y), \ldots, U_{m|N|-1}(Y)\right]\right\}, \tag{1}$$

where

$$U_{m1}(Y) \leq U_{m2}(Y) \leq \cdots \leq U_{m|N|-1}(Y). \tag{2}$$

(i) Apart from the root, each node has at least one outgoing edge:

$$\sum_n y_{mn} \geq 1 \quad \text{for each } m \in N,\ m \neq r. \tag{3}$$

(ii) All edges incoming to the root are included in the DODAG:

$$y_{nr} = 1 \quad \text{for each neighbour } n \in N. \tag{4}$$

(iii) At the same time, all edges outgoing from the root are excluded from the DODAG:

$$y_{rn} = 0 \quad \text{for each neighbour } n \in N. \tag{5}$$

Note 1. Instead of constraints (4) and (5) and variables y_{nr} and y_{rn}, binary constants might be inserted.

(iv) Of each edge pair (m, n) and (n, m) not adjacent to the root, not more than one should be selected:

$$y_{mn} + y_{nm} \leq 1 \quad \text{for each pair of neighbouring nodes } m, n \in N,\ m \neq r,\ n \neq r. \tag{6}$$

Note 2. In the case of maximizing the number of paths, there will always be one direction selected; that is, the constraint will always be met with equality.

(v) All possible cycles have to be eliminated. Cycles of length 1 do not exist in the network graph, as it is assumed to be simple. Cycles of length 2 are already eliminated by constraints (6). Hence, it is required to eliminate cycles of length $K = 3, 4, \ldots, |N|$. For this purpose, the following constraints can be formulated:

$$y_{k1k2} + y_{k2k3} + \cdots + y_{kKk1} \leq K - 1$$
$$\text{for each } K \text{ interconnected nodes, } k_1, k_2, \ldots, k_K \in N. \tag{7}$$

(vi) If any of the edge variables belonging to a path p is equal to 0, then the variable Y_p is also equal to 0, which gives $Cp - 1$ constraints (8) per path (not C_p as, due to constraints (4), edges incoming to the root are always included in the DODAG):

$$Y_p \leq y_{mn} \quad \text{for each pair of consecutive nodes } m, n \in N \text{ on the path } p, \text{except } r\ (n \neq r). \tag{8}$$

On the other hand, if a path p is included in the DODAG due to maximization ($Y_p = 1$), then all of its edges will be included in the DODAG; that is, all y_{mn} variables of this path will be equal to 1.

(vii) The last expression represents the number of paths from node m to the root:

$$U_m = \sum_p Y_p \quad \text{for each } m \neq r \text{ and paths } p \text{ originating at node } m. \tag{9}$$

2.2. Proof of Correctness. Constraints presented in the previous subsection are sufficient to construct a valid DODAG due to the following facts:

(i) all cycles are eliminated due to constraints (6) and (7);

(ii) each node has at least one path to the root. This path does not necessarily belong to the candidate paths set inserted into the formulation. It might be one of the additional paths resulting from the edge orientation of the solution. This requirement basically means that no other root than the one given is created in the resulting graph; therefore, it is a DODAG. This can be proved in the following way: due to the constraints (3), any arbitrary node different than the root has at least one outgoing edge. Following one of the outgoing edges, it is possible to reach either the root or another node, which also has at least one outgoing edge. Due to the fact that no cycles exist, which means that it is impossible to come back to any of the nodes traversed previously, and the fact that by applying constraints (3) and (5) the root is the only node with no outgoing edges, the whole process can finish only at the root.

2.3. The Lexicographic Algorithm. The presented DODAG construction problem belongs to the class of equitable resource allocation problems with integer decisions, where the number of possible distinct outcomes of the performance functions is limited and relatively small. Here, a single performance function, that is, a $U_{mi}(Y)$ function representing the number of paths of a single node, might assume all integer values from 0 to k_m, where k_m is the number of candidate paths inserted into the formulation for this node. Hence, the set of possible distinct outcomes of all performance functions contains all integer values between 0 and $k_{\max}$, where $k_{\max}$ is the largest number of candidate paths inserted for any node.

An algorithm designed for solving this class of problems is presented in [11], Section 7.2.3 *Lexicographic Minimization of Counting Functions*. The algorithm in its original form solves the problem of lexicographic minimization. The number of performance functions' possible distinct outcomes determines the maximum number of lexicographic iterations. The basic idea of this algorithm is the following. The set of all possible distinct outcomes has to be provided. Special counting functions are then constructed, each of them representing the number of times that a single distinct outcome appears in the solution. These functions are then iteratively minimized in the following way. First, the number of occurrences of the largest possible outcome is minimized. Then, the number of occurrences of the second largest outcome is minimized without increasing the first one and so forth. The algorithm terminates after reaching the last (smallest) distinct outcome or in an earlier iteration, if a unique solution is found. Due to the fact that the counting functions are represented as optimization problems, additional constraints and continuous variables have to be inserted in each lexicographic iteration, extending the problem and transforming it from an IP to a MIP problem.

The lexicographic maximization objective can be transformed into lexicographic minimization objective, similar to the maximin-minimax conversion given in Section 1.2 of [11]:

$$\begin{aligned} &\operatorname{lexmax}\left\{U^{(m)}(Y)=\left[U_{m1}(Y), U_{m2}(Y), \ldots, U_{m|N|-1}(Y)\right]\right\} \\ &\quad=-\operatorname{lexmin}\left\{-U^{(m)}(Y)\right. \\ &\qquad\qquad\left.=\left[-U_{m1}(Y),-U_{m2}(Y), \ldots,-U_{m|N|-1}(Y)\right]\right\}, \end{aligned} \tag{10}$$

where

$$-U_{m1}(Y) \geq -U_{m2}(Y) \geq \cdots \geq -U_{m|N|-1}(Y). \tag{11}$$

2.4. The Exact Number of Distinct Outcomes. As mentioned already, the set containing all integers between 0 and $k_{\max}$ can be used as the set of all possible distinct outcomes of the performance functions in this formulation. This method is approximate and results in an overhead, since it is very likely that two or more paths inserted into the formulation for a single node will exclude each other from the resulting DODAG because of using one or more edges of opposite direction between the same pair of nodes. To obtain the minimal upper bound on the possible distinct outcomes set, a maximum independent set count problem would have to be solved for every node, as proved in Appendix A, and the maximum of the obtained values should be used as the upper bound. However, due to the NP-hardness of the maximum independent set problem, it might be of advantage to omit the exact computation of the maximum distinct outcome value, even though a crude approximation increases the maximum number of lexicographic iterations.

2.5. Size of the Formulation: Reduction of the Number of Cycle Elimination Constraints. The cycle elimination constraints (7), due to their numerosity, incur heavy computational load while solving the formulation. Maximally $2\cdot(\binom{N}{3}+\binom{N}{4}+\cdots+\binom{N}{N})$ (in the case of a complete graph) constraints (7) could appear in the problem. Although on average this number will be smaller, the number of cycles in a graph, *ergo* the number of constraints, grows exponentially as a function of the size of the graph (in terms of edges). To reduce the number of inserted cycle elimination constraints, the following method has been employed (based on a suggestion of A. Tomaszewski):

(i) the formulation is solved without constraints (7);

(ii) if cycles exist in the resulting graph, appropriate (7) cycle elimination constraints are inserted. Only a few cycles are searched for (e.g., up to 5);

(iii) the extended formulation is solved again, until the output graph is a DODAG; that is, it does not contain any cycles.

This method trades solving one large IP problem for iterative solving of several smaller IP problems. In the last of the solved problems, all possible cycles will be eliminated, yet avoiding inclusion of unnecessary constraints. When solving the whole lexicographic formulation, all cycle elimination constraints added in the ith iteration are transferred to the $i+1$th lexicographic iteration.

3. Simulations and Results

3.1. Generated Networks. Random *ad hoc* networks were generated for the simulations. WPA *ad hoc* network generation algorithm presented in [12] was used, with a small modification and with values of the parameters chosen so that generated networks were always connected and of reasonable density (limiting the density limits the number of existing paths in the network graphs). All values for parameters used in the WPA algorithm as well as a short remark about the modification are given in Appendix C.

The output of the WPA is an undirected graph, where the length of each edge represents the distance between the two nodes placed on a 2D surface. In the simulations, each undirected edge was replaced by a pair of inverse directed edges. For one of the tested algorithms, a weight was assigned to each directed edge, representing 2.45 GHz signal attenuation in free space (This frequency is used, for example, in Zigbee technology.). The weight was calculated according to (7) in [13] and length of the edge used as the distance, assuming meters as units. A random deviation of ±10% was also added to each link weight.

Networks comprising 5 to 50 nodes were generated, with a step of 5. For each network size, 50 random networks were generated. The first generated node in the network was always selected as the root node.

3.2. Input Paths. Sets of k shortest paths, one set per node other than the root, were inserted into the formulation, where *shortest* means of least hop count. Yen's k shortest path algorithm was used for calculating these sets [14]. BFS algorithm was used as the shortest path subalgorithm. k values were equal for every node in the network. They were

Table 1: Comparison of four tested algorithms: average number of paths per node at different levels 1–6 for 50-node networks. Levels (1–6) were present in most (47–50 out of 50) networks.

Level	Shortest-multipath	TORA	MARA-MC	k-shortest-lex-maximin
1	2.17	2.25	3.22	3.72
2	4.67	4.39	5.50	5.85
3	7.00	6.32	6.97	7.06
4	8.57	7.69	7.66	8.00
5	9.65	8.42	8.04	8.72
6	10.46	9.44	8.55	9.43

Table 2: Comparison of four tested algorithms: average variance in numbers of paths per node at different levels 1–6 for 50-node networks.

Level	Shortest-multipath	TORA	MARA-MC	k-shortest-lex-maximin
1	1.64	2.24	4.12	4.08
2	5.77	5.27	4.28	2.56
3	9.02	8.36	4.65	2.59
4	9.25	8.29	6.12	3.01
5	10.29	11.93	7.72	3.70
6	9.93	10.83	11.39	5.32

dependent on the size of the network and equal to 5 for networks of sizes 5 to 10, 10 for 15 to 30-node networks, and 15 for 35 to 50-node networks.

3.3. Tests. DODAGs formed by three distributed algorithms that appear in the literature were compared against the k-shortest-lexicographic solution: the Shortest-multipath DODAG as in MDVA [2], using link costs as given in Section 3.1, the MARA-MC DODAG [7], and a DODAG approximating the output graph of the TORA *ad hoc* multipath routing algorithm [6]. The approximation mentioned above avoids random irregularities specific to the graphs constructed by the TORA algorithm. Hence, to form a DODAG, the nodes are simply ordered by the hop count metric (edges are directed from the node with the greater hop count to the node with the smaller hop count), which, in fact, makes the resulting DODAG similar to a Shortest-multipath DODAG when link costs are equal to 1. Ties in hop counts are resolved using nodes' unique identifiers. This simplification might incur slightly different results than TORA would achieve in reality, but it is required to provide the possibility of performing the simulations in a simple way (i.e., without employing more complex network simulators).

An especially interesting algorithm is MARA-MC [7], which solves an all-to-one maximum edge connectivity problem with optimality. The authors' simulations show that this algorithm obtains significantly more paths in the resulting DAGs than other compared algorithms and that it calculates a large number of paths for a large fraction of source-destination pairs. It is therefore interesting to check how well it performs in comparison to the lexicographic algorithm. It is worth noting that the all-to-one maximum connectivity objective as defined in Section 4.A of their paper does not have any meaning in the case of nonmultigraphs, as the minimum connectivity between a node and a simple DAG root will always be equal to 1, regardless of the algorithm used. This is an implication of Proposition 1. A possibly significant drawback of the MARA-MC algorithm is that it completely ignores the length of the resulting paths.

3.4. Results. Tables 1 and 2 present the average value and the average variance in the number of candidate (shortest) paths per node at different levels, in the DODAGs constructed by the three algorithms from the literature and by the lexicographic algorithm, when testing 50-node networks. At level 1, one node with the number of paths equal to 1 was excluded from the calculations, since, due to Proposition 1, it is always present, regardless of the algorithm used (as its only path is the shortest one (hop count = 1), it is always included in the input path set).

While for levels 2–6 no substantial differences in the average values can be observed, Table 2 shows that for these levels the equitable solution achieves significantly smaller variance values than the three other algorithms. These two observations mean that, for these levels, the numbers of shortest paths of different nodes in the equitable solution are much more balanced. The closest to it is the MARA-MC algorithm, which has also outperformed the other two algorithms considerably as regards variance values at levels 2–5. At level 1, the relation between the variance values and average values is similar for all the algorithms, although the equitable solution and MARA-MC obtain larger average numbers of paths.

Another interesting observation is that the Shortest-multipath algorithm, with the link costs calculated basing on the 2D distance, obtains greater average numbers of shortest hop count paths than the TORA approximation algorithm, which is based on the hop count metric. This happens probably because, in the simulated Shortest-multipath, a node of a smaller cost would most likely have paths of smaller hop count lengths than a node of a greater cost. In the case of the TORA algorithm, no additional information is available

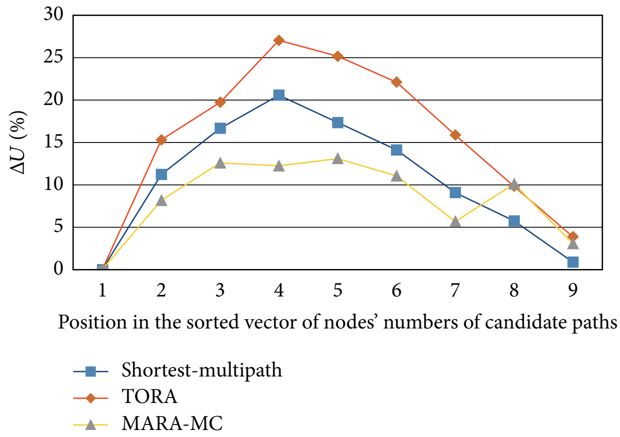


FIGURE 2: Relative comparison of the three algorithms for 10-node networks using the lexicographic k-shortest solution as a reference. ΔU values are relative to the reference (lexicographic) solution.

to grade the nodes at the same hop count distance from the root; hence, a significant fraction of edges is assigned at random (using node identifiers).

Figure 2 shows a comparison of the three algorithms on 10-node networks, with the lexicographic k-shortest solution used as a reference. The ΔU value reflects the average gap to the reference solution on individual lexicographic positions and is defined as follows:

(i) let U_{i_ref} = the number of paths (only taking into account the input k shortest paths) in a given position in the equitable solution vector of the ith network, and let U_{i_alg} = the number of paths in the same position in the vector obtained by the tested algorithm for the ith network (previously sorted in a nondecreasing order so that a lexicographic comparison using individual positions is possible), $i = 1, 2, \ldots, 50$;

(ii) let $\Delta U_i = ((U_{i\_\mathrm{ref}} - U_{i\_\mathrm{alg}})/U_{i\_\mathrm{ref}}) \times 100[\%]$;

(iii) then, $\Delta U = (\sum_i \Delta U_i)/50$.

Figure 2 shows, again, that the MARA-MC algorithm is the best of the three algorithms in terms of equitability in the distribution of shortest paths, as it has, on average, vectors whose values in the lowest positions are closest to the equitable solution vectors.

An interesting observation can be made that all algorithms achieve a 0% gap in the first lexicographic position. This is due to the fact mentioned previously, namely, that one of the root's neighbours in every simple DODAG will always have exactly one path to the root, with this path being one of this node's shortest paths (hop count = 1). This node probably always occupies the first lexicographic position in the case of 10-node networks.

However, this is clearly not always the case when dealing with 50-node networks, as shown in Figure 3. It can be observed that the MARA-MC algorithm sometimes happens to have a nonzero gap in the first position, which means that, in some cases, there exists at least one node in the DODAG obtained by MARA-MC which has no paths available from

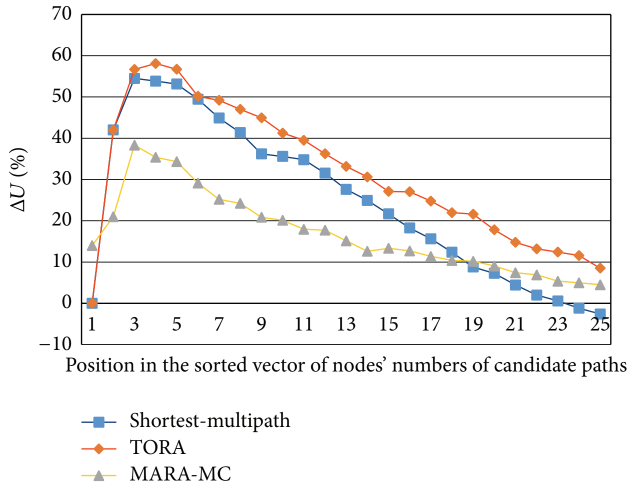


FIGURE 3: Relative comparison of the three algorithms for 50-node networks using the k-shortest lexicographic solution as a reference: first 25 positions in the sorted solution vectors.

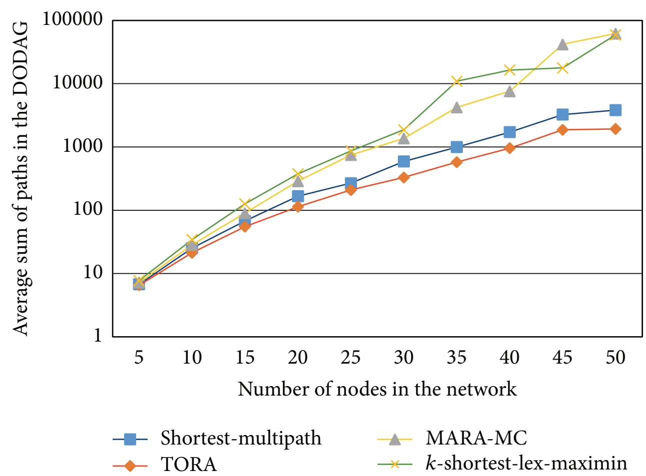


FIGURE 4: Average sum of all paths in the DODAGs.

the set inserted into the formulation for that node—which means that this node has only longer paths than the $k = 15$ shortest possible paths and occupies the first lexicographic position instead of the root's 1-path neighbour. Moreover, although MARA-MC performs better in the overall comparison to the two other algorithms, it is still far from equitability achieved by the lexicographic algorithm (~20%–38% gaps in positions 2–10).

Figure 4 shows the average sums of all paths in the obtained DODAGs; this time, all possible paths were enumerated (not only the candidate paths from the shortest path sets inserted into the lexicographic formulation but also the additional paths resulting from the obtained edge orientation). It can be concluded from the plot that the most balanced solutions, like the k-shortest lexicographic solution and the MARA-MC solution, also achieve the greatest numbers of paths in general.

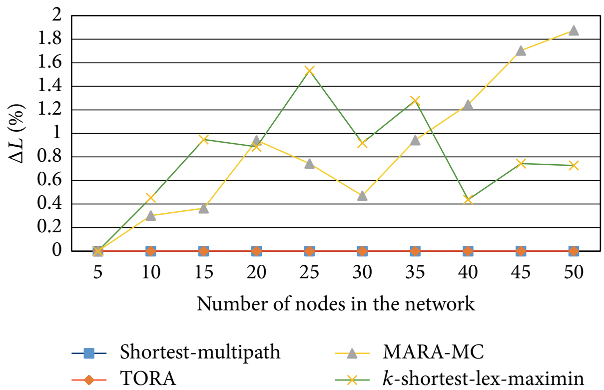


FIGURE 5: Average DODAG-to-network shortest path length gap in function of the size of the network.

Figure 5 shows the average gap between the hop count length of the shortest path of a single node available in the formed DODAG and the length of this node's actual shortest path, that is, this node's shortest path in the network. Hence, the ΔL value is defined as follows:

(i) let $L_{\text{in\_opt}}$ = the length of the shortest path of a node n in ith network, and let $L_{\text{in\_alg}}$ = the length of the shortest path for this node in the DODAG obtained by the algorithm;

(ii) let $\Delta L_{\text{in}} = ((L_{\text{in\_alg}} - L_{\text{in\_opt}})/L_{\text{in\_opt}}) \times 100[\%]$;

(iii) hence, $\Delta L = (\sum_i \sum_n \Delta L_{\text{in}})/(50(N-1))$, where $N - 1$ = the number of nodes in the network excluding the root.

An interesting observation can be made, namely, both MARA-MC and the lexicographic k-shortest maximin algorithm do not achieve zero ΔL values for networks greater than 5 nodes, which means that the output DODAGs do not always include the whole shortest path trees. This in turn means that to achieve fairness in terms of the numbers of input paths per node in a DODAG, even when considering sets of shortest paths as it has been done in the simulations, a trade-off has to be made between the equitability and availability of the shortest paths of some nodes.

4. Conclusions

A lexicographic optimization formulation has been proposed in this paper, solution of which achieves an equitable distribution of the paths available per node in a constructed DODAG. This property could be important in the context of multipath routing, where network topology changes are infrequent. The formulation allows insertion of the user's own sets of paths per node. This property makes the mechanism adjustable to particular requirements imposed by the application on the structure of paths, although a well-thought-out choice of the input path sets is necessary to obtain reasonable results.

A comparison of three distributed algorithms that appear in the literature: MARA-MC, Shortest-multipath, and TORA, and the lexicographic algorithm has been presented, with k shortest paths per node being inserted into the formulation. The lexicographic solution has been proved to obtain the best results in terms of providing a fair distribution of the shortest paths of the nodes at the same minimum hop count distance from the root (at the same level).

Two interesting properties of the k-shortest equitable solution have been observed. First, it has been shown that providing efficient maximin fairness in the numbers of paths per node increases the overall number of paths in the resulting DODAG. Second, even though sets of shortest paths were included in the optimization problem, the resulting DODAGs did not always contain full shortest path trees.

It has been shown that of the three distributed algorithms, MARA-MC achieves the best results, although it still does not get close to the k-shortest equitable solution in low lexicographic positions. Moreover, it turns out that, in some cases of 50-node networks, a node could be found in the MARA-MC DODAG whose shortest path available in the graph was longer than its first 15 shortest paths in the network. These observations lead to a conclusion that designing a distributed algorithm to construct a *fair* DODAG, that is, with equitable distribution of the paths per node, is still an open question.

Appendices

A. Proof of NP-Completeness of the Simplified Problem

The simplified problem is stated as follows. Given a simple directed graph $G = (N, E)$, two nodes $s, t \in N$ and a list of paths $p_1, p_2, \ldots, p_P \in P$, from node s to node t, find a subset of edges such that no two edges between the same pair of nodes are included (i.e., no edges of conflicting directions are included) and the number of available paths between s and t is maximized. Let the simplified problem be called MAX-PATHS.

It is possible that two or more paths from the P set will exclude each other because of using one or more edges of opposite directions between the same pair of nodes. For example, if, for the network shown in Figure 6, the following three paths belong to the P set: (1) s-a-t, (2) s-a-b-t and (3) s-b-a-t, then the maximum possible number of paths from s to t for the given set will be equal to 2, as paths s-a-b-t and s-b-a-t cannot both be included due to the conflict in assignment of the direction of edge ab.

There might be more than two counter-direction paths. This problem can be modeled as shown in Figure 7. Vertices $p_1 - p_5$ represent an example set of paths between s and t. An edge exists between a pair of vertices if the corresponding paths cannot be picked together due to at least one edge conflict. Therefore, the maximum number of paths from node s to t that can be chosen together is equal to the size of the maximum independent set of this graph. The vertices in Figure 7 that belong to the maximum independent set of the graph have been marked green. Hence, in this case, the objective value is equal to 3, although 5 paths were considered originally.

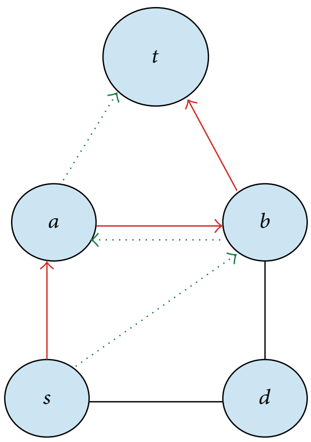


Figure 6: If for node s, paths s-a-b-t (red solid) and s-b-a-t (green dotted) are given, then these paths will not both be included due to the conflict in the used direction of edge ab.

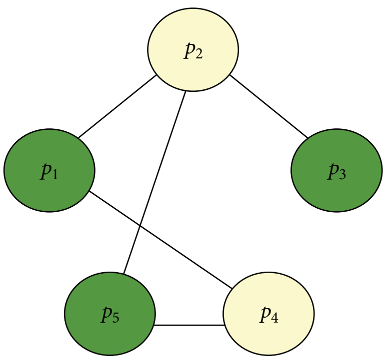


Figure 7: An edge exists between a pair of vertices, if the corresponding paths cannot coexist. For example, path p_3 cannot exist together with path p_2. Paths belonging to the maximum independent set have been marked green. In this case, the maximum set of paths that can be picked together is equal to 3.

Therefore, to show that the MAX-PATHS problem is NP-complete, instances of the maximum independent set (MIS) problem can be mapped to the MAX-PATHS problem. It can be shown that a valid solution to MIS is also a valid solution to MAX-PATHS problem and that a nonvalid solution to MIS is not a valid solution to MAX-PATHS.

Consider an MIS problem instance, modeled with a graph $G'(N', E')$. A corresponding instance of the MAX-PATHS problem, modeled by a network graph $G(N, E)$ and the set P of paths, can be constructed in the following way.

(1) For every edge $e' \in E'$ adjacent to a pair of nodes $n'_1, n'_2 \in N'$ in the MIS problem, create two vertices $n_{e'1}, n_{e'2} \in N$ in the MAX_PATHS problem.

(2) In MAX_PATHS, between each pair $n_{e'1}$, $n_{e'2}$, add two edges $e_1, e_2 \in E$ of opposite directions, hence obtaining a bipartite graph.

(3) In each MAX_PATHS pair $n_{e'1}$, $n_{e'2}$, label one of the e_1, e_2 edges with node n'_1 of the corresponding edge e' in MIS, and the other with the node n'_2.

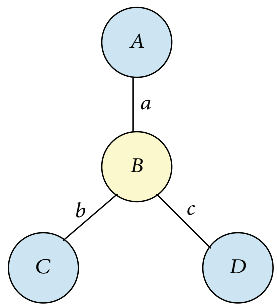


Figure 8: G' graph of the MIS problem. Vertices A, C, and D belong to the maximum independent set.

(4) In MAX_PATHS, add the source and destination nodes $s, t \in N$.

(5) For each node $n' \in N'$ in MIS, build a path $p_{n'} \in P$ from s to t in MAX_PATHS, going through all the edges labeled with n', and adding necessary edges $e \in E$: between s and the $N \setminus \{s, t\}$ set, between the $N \setminus \{s, t\}$ set and t, and between the nodes in the $N \setminus \{s, t\}$ set, belonging to different pairs $n_{e'1}, n_{e'2}$. The order of constructing the paths is arbitrary.

Proposition 2. *An independent set in G' corresponds to a set of nonconflicting paths in G.*

In the MAX_PATHS problem, there is one path $p_{n'} \in P$ from s to t for each node $n' \in N'$ in MIS. Each of these paths goes through exactly $2M_{n'}$ nodes, where $M_{n'}$ is the number of n''s neighbours in G'. More precisely, each path crosses $M_{n'}$ pairs of $n_{e'1}$, $n_{e'2}$. Every time a path is selected, an edge orientation between $n_{e'1}$ and $n_{e'2}$ is fixed, forbidding the selection of a conflicting path.

(6) Notice that if an independent set on G' corresponds to a set of paths that do not have conflicting edges in G, the maximum independent set in MIS will correspond to the maximum number of nonconflicting paths from s to t in MAX_PATHS.

The mapping process has been shown in Figures 8, 9, 10, and 11.

B. Proof of Proposition 1

Proposition 1 is proved as follows:

(i) consider a network comprising a single node which is the root of the DODAG;

(ii) add a neighbouring node A with a single edge directed to the root. Node A has exactly one path to the root. To increase node A's number of paths to the root without cycles, another node has to be added that has a connection both to the root (but not via node A) and from node A;

(iii) add a node B, with an edge directed to the root and an edge incoming from node A. Node A has now two

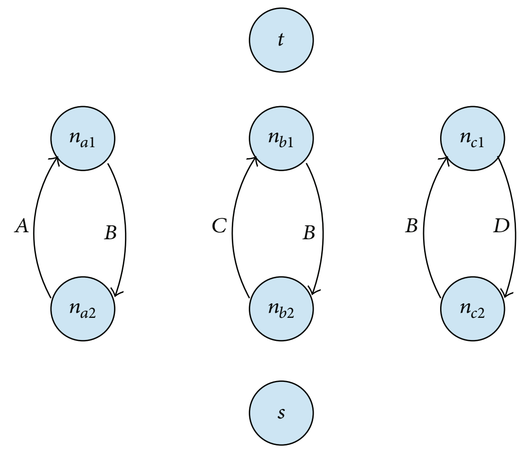


Figure 9: G mapping after step 4.

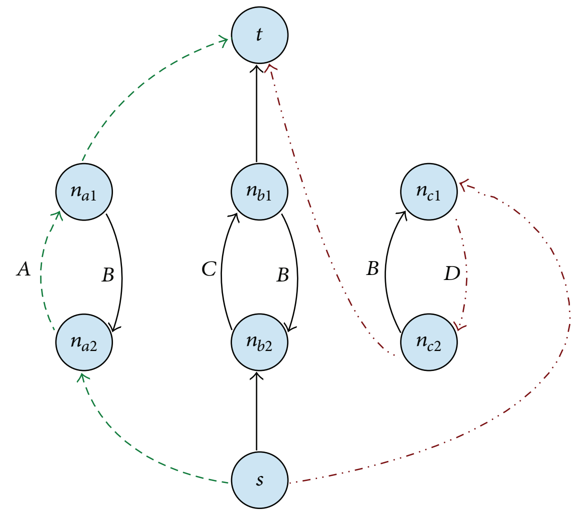


Figure 10: A, C, and D paths belonging to the maximum independent set. These paths do not have any conflicting edges.

paths to the root; however, node B is now the one that has only one path to the root;

(iv) adding an edge from node B to node A will result in a cycle; therefore, the only way to provide node B with a second path is to connect it to another node, which has either a direct link to the root (is another neighbour of the root) or has a path to the root which goes through another neighbour of the root. This neighbour will, again, either have only one path to the root or its other paths will eventually lead to another root's neighbour with this property.

The proof is shown in Figures 12, 13, and 14.

The proof of this proposition implicates that

(i) the minimum number of paths of a single node to the root in a simple DODAG is always equal to one;

(ii) the minimum connectivity between a single node and the root in a simple DODAG is always equal to one;

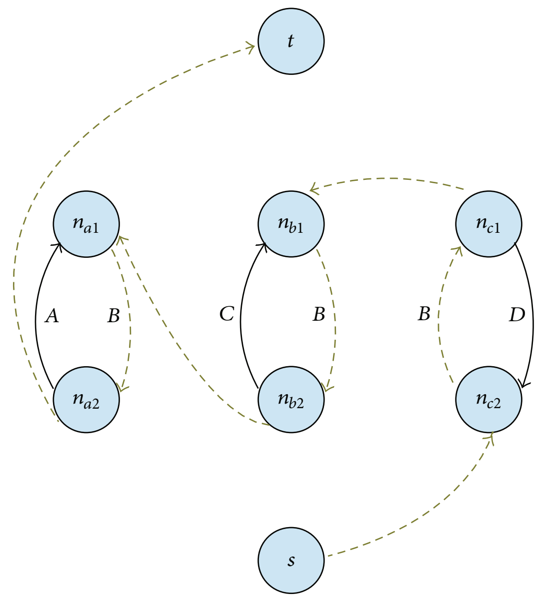


Figure 11: B path, conflicting with the paths from the maximum independent set due to different selection of edge directions.

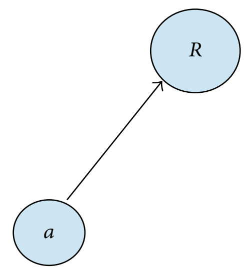


Figure 12: Node a has only one path to the root.

Table 3: Parameter values for the WPA algorithm.

Parameter	Value
r	12
d_0	6
l	100

(iii) this one path is the shortest path of the node (hop count = 1).

C. Parameter Values for the WPA Algorithm

Table 3 contains parameter values for the WPA algorithm. The same values were used for networks of different sizes.

Change in the Generation Procedure. Instead of calculating the transmission range after generating node positions according to the length of resulting edges, it was assumed to be equal to r to ensure connectivity and prevent repeated generation of node positions.

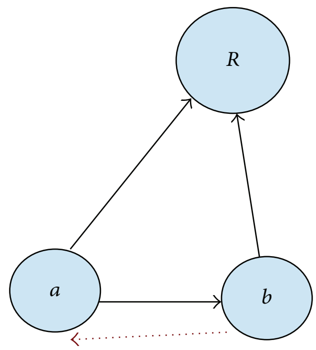


FIGURE 13: Node a has now two paths to the root, but node b has just one. Creating a second path for node b by adding an edge directed to node a will create a cycle.

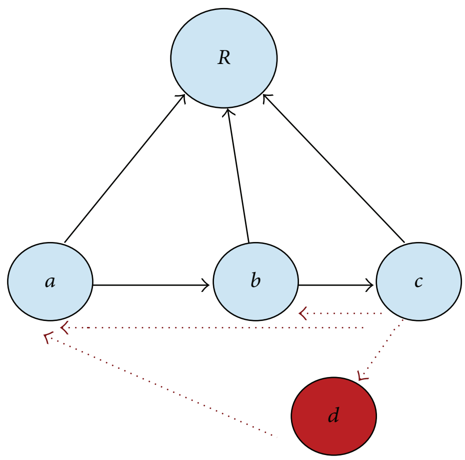


FIGURE 14: A similar situation happens when a third node is added. And so on.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank Dr. Artur Tomaszewski for suggesting the iterative method of simplifying the formulation by reducing the number of cycle elimination constraints, mentioned in Section 2.5 of this paper, and Dr. Hanan Luss for a small but valuable hint and his kind willingness to help. The research presented in this paper was partially supported by the National Science Centre (Poland) under Grant 2011/01/B/ST7/02967 "Integer programming models for joint optimization of link capacity assignment, transmission scheduling, and routing in fair multicommodity flow networks." The research presented in this paper has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under Grant Agreement no. 269985.

References

[1] T. Winter, P. Thubert, A. Brandt et al., "RPL: IPv6 routing protocol for low-power and lossy networks," Internet Engineering Task Force RFC 6550, 2012.

[2] S. Vutukury and J. J. Garcia-Luna-Aceves, "MDVA: a distance-vector multipath routing protocol," in *Proceedings of the 20th Annual Joint Conference on the IEEE Computer and Communications Societies (IEEE INFOCOM '01)*, vol. 1, pp. 557–564, Anchorage, Alaska, USA, April 2001.

[3] S. Vutukury and J. J. Garcia-Luna-Aceves, "An algorithm for multipath computation using distance-vectors with predecessor information," in *Proceedings of 8th IEEE International Conference on Computer Communications and Networks*, pp. 534–539, Boston, Mass, USA, 1999.

[4] S. Vutukury and J. J. Garcia-Luna-Aceves, "A simple approximation to minimum-delay routing," in *Proceedings of the ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (SIGCOMM '99)*, pp. 227–238, Cambridge, Mass, USA, 1999.

[5] M. S. Corson and A. Ephremides, "A distributed routing algorithm for mobile wireless networks," *Wireless Networks*, vol. 1, no. 1, pp. 61–81, 1995.

[6] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *Proceedings of the 16th Annual Joint Conference of the IEEE Computer and Communications Societies. Driving the Information Revolution (IEEE INFOCOM '97)*, p. 1405, Kobe, Japan, April 1997.

[7] Y. Ohara, S. Imahori, and R. van Meter, "MARA: maximum alternative routing algorithm," in *Proceedings of the 28th Conference on Computer Communications (IEEE INFOCOM '09)*, pp. 298–306, Rio de Janeiro, Brazil, April 2009.

[8] S. Cho, T. Elhourani, and S. Ramasubramanian, "Independent directed acyclic graphs for resilient multipath routing," *IEEE/ACM Transactions on Networking*, vol. 20, no. 1, pp. 153–162, 2012.

[9] M. Pióro and D. Medhi, "Chapter 5, section 5.5: gradient minimization and other approaches for convex programming problems," in *Routing, Flow and Capacity Design in Communication and Computer Networks*, Morgan Kaufmann, Boston, Mass, USA, 2004.

[10] K. Zhang and H. Gao, "Finding multiple length-bounded disjoint paths in wireless sensor networks," *Wireless Sensor Network*, vol. 3, no. 12, pp. 384–390, 2011.

[11] H. Luss, "Chapter 1: introduction. Chapter 2: nonlinear resource allocation. Chapter 7: equitable resource allocation with integer decisions, section 7.2: problems with a limited number of distinct outcomes," in *Equitable Resource Allocation: Models, Algorithms, and Applications*, John Wiley & Sons, New York, NY, USA.

[12] F. A. Onat, I. Stojmenovic, and H. Yanikomeroglu, "Generating random graphs for the simulation of wireless ad hoc, actuator, sensor, and internet networks," *Pervasive and Mobile Computing*, vol. 4, no. 5, pp. 597–615, 2008.

[13] S. Wloczysiak, "Extending 2.4 GHz ZigBee short range radio performance with skyworks front-end modules: RF signal propagation attenuation," *Microwave Journal*, 2009.

[14] J. Y. Yen, "Finding the K shortest loopless paths in a network," *Management Science*, vol. 17, no. 11, pp. 712–716, 1971.

Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2014, Article ID 469437, 8 pages
http://dx.doi.org/10.1155/2014/469437

Research Article

Optimization of Power Allocation for a Multibeam Satellite Communication System with Interbeam Interference

Heng Wang, Aijun Liu, Xiaofei Pan, and Jiong Li

College of Communications Engineering, PLA University of Science and Technology, No. 2 , Yudao Street, Nanjing 210007, China

Correspondence should be addressed to Heng Wang; wangheng0987654321@126.com

Received 22 October 2013; Accepted 24 December 2013; Published 16 January 2014

Academic Editor: Hanan Luss

In multibeam satellite communication systems it is important to improve the utilization efficiency of the power resources, due to the scarcity of satellite power resources. The interbeam interference between the beams must be considered in the power allocation; therefore, it is important to optimize the power allocated to each beam in order to improve the total system performance. Initially the power allocation problem is formulated as a nonlinear optimization, considering a compromise between the maximization of total system capacity and the fairness of the power allocation amongst the beams. A dynamic power allocation algorithm based on duality theory is then proposed to obtain a locally optimal solution for the optimization problem. Compared with traditional power allocation algorithms, this proposed dynamic power allocation algorithm improves the fairness of the power allocation amongst the beams, and, in addition, the proposed algorithm also increases the total system capacity in certain scenarios.

1. Introduction

In satellite communication systems, a satellite may provide coverage of the entire region of the earth visible from the satellite, by using a single beam. In this case, the gain of the satellite antenna will be limited by the beamwidth, as imposed by the coverage. For instance, for a geostationary satellite, global coverage implies a 3 dB beamwidth of 17.5 and consequently an antenna gain of no more than 20 dB [1]. Therefore, each user must be equipped with a large aperture antenna to support the high traffic rate, which results in great inconvenience. In order to solve this problem, the multibeam technique has been widely applied in modern satellite communication systems. In multibeam satellite communication systems, the satellite provides coverage of only part of the earth, by means of a narrow beam. The benefit of a higher satellite antenna gain is obtained due to a reduction in the aperture angle of the antenna beam [1]. As a result, a user with a small aperture antenna can support a high traffic rate. Moreover, the multibeam technique supports the reuse of frequencies for different beams, in order to increase the total system capacity. When two beams utilize the same frequency, interbeam interference is introduced to the two beams, due to the nonzero gain of the antenna side lobe. It has been noted that when there is interbeam interference between the beams, the capacity allocated to each beam is determined not only by the power allocated to the beam, but also by the power allocated to the other beams.

Due to the limitations of satellite platform, it is known that satellite power resources are scarce and expensive. It is thus important to optimize the utilization efficiency of the power resources. Moreover, the traffic demands of each beam are different, with varying times, due to the different coverage areas, and the interbeam interference between the different beams is also different. As a result, it is critical to optimize the power allocation to each beam to meet the specific traffic demands.

Power allocation algorithms were proposed in earlier works [2–7]. The mathematical formulation and analytic solutions of the optimum power allocation problem have been presented [2]; however, the mathematical algorithm to solve the optimization problem was not provided. As a result, bisection and subgradient methodologies have been utilized to solve the optimization problem [3, 4]. In order to improve the total system capacity, a method to select a small number of active beams has been proposed [5], which maintained the fairness of the power allocation amongst the beams. The main problem in [2–5] was that the authors failed to consider the interbeam interference between the beams, which cannot be ignored in determining power allocations. A novel resource allocation scheme for multibeam satellite communication systems has been described, offered maximum communication capacity [6]. The scheme optimized the frequency bandwidth, the satellite transmission power, the modulation level, and the coding rate to each beam, in order to manage the ever-changing user distributions and the interbeam interference conditions. However, the scheme ignored the fairness of the power allocations amongst the beams. A joint optimization allocation algorithm for the power and the carrier was proposed [7], in order to best match the asymmetric traffic requests. The algorithm attempted to support the greatest degree of fairness in the power allocation to each beam, regardless of the total system capacity.

This paper's research is aimed at resolving this deficiency, by optimizing the power allocations for a multibeam satellite communication system, with full consideration of the impact of interbeam interference. The first step is to mathematically formulate the problem of power allocation as a non-linear optimization, compromising between the maximization of total system capacity and the fairness of the power allocations to each beam. It is found that, in the optimization process, the optimal variables are coupled with each other. As a result, it is difficult to determine whether the optimization is convex or not and to obtain the globally optimal solution for the optimization. To this end, a dynamic power allocation algorithm based on duality theory is proposed to obtain a locally optimal solution for the optimization. Finally, the simulation results show the efficiency of the proposed dynamic power allocation algorithm.

The main contributions of this research are summarized as follows:

(1) the mathematical formulation of the power allocation problem for multibeam satellite communication, with consideration of interbeam interference, through a compromise between the maximization of the total system capacity and the fairness of the power allocation amongst the beams;

(2) the proposal of an algorithm, based on duality theory, which will obtain a locally optimal solution for the optimization problem;

(3) a demonstration of the effects of the interbeam interference and the channel conditions of each beam on the power allocation results.

The remainder of this paper is organized as follows. In Section 2, the model of a multibeam satellite communication system with interbeam interference is described. In Section 3, a mathematical formulation of the optimization problem of the power allocation is presented. Section 4 presents the proposal of a dynamic power allocation algorithm designed to obtain a locally optimal solution to the optimization. Section 5 presents the simulation results and analyzes the effects of the interbeam interference and channel conditions of each beam on the power allocation results. Section 6 presents the conclusion of the paper.

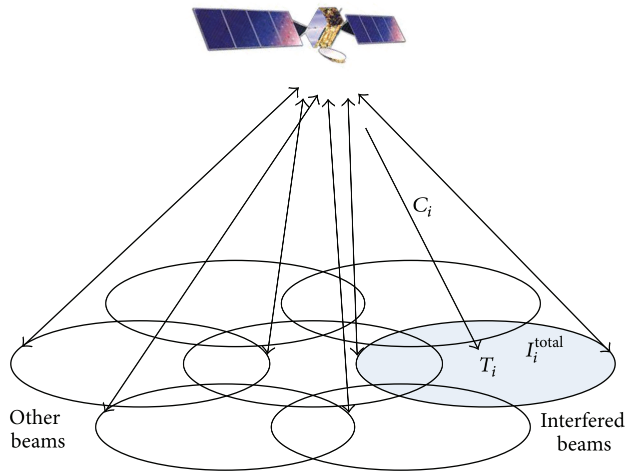


Figure 1: Configuration of the multibeam satellite communication system.

2. A Multibeam Satellite Communication System Model

Figure 1 shows the system configuration of the multibeam satellite communication system that is studied here, where

N: is the quantity of the beams,

T_i is the traffic demand of the ith beam,

P_i is the power allocated to the ith beam,

I_i^{total} is the total interference on the ith beam from the other beams,

γ is the signal attenuation factor of the ith beam, and it is noted that γ mainly consisted of the effects of weather conditions, free space loss, and antenna gain, and

P_{total} is the total satellite power resources within the system.

To precisely describe the interbeam interference within the system, the interbeam interference matrix H is introduced, which is defined as follows:

$$H = \begin{bmatrix} 0 & h_{12} & \cdots & h_{1n} \\ h_{21} & 0 & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & 0 \end{bmatrix}, \quad (1)$$

where the element h_{ij} denotes the interference coefficient from the jth beam on the ith beam. It is noted that the element h_{ii} is zero, because the interference from the same beam is ignored. It is obvious from (1) that the total interference on the

ith beam from other beams I_i^{total} is $\sum_{k=1,k\neq i}^{N} P_k h_{i,k}$. As a result, using time sharing for Gaussian broadcast channels [8], the Shannon bounded capacity C_i for the ith beam is given as

$$C_i = W \log_2 \left(1 + \frac{P_i}{\gamma W N_0 + \sum_{k=1,k\neq i}^{N} P_k h_{i,k}}\right), \quad (2)$$

where N_0 is the noise power density of each beam and W is the bandwidth of each beam. It is shown in (2) that the capacity C_i of the ith beam is increased as the power allocated to the beam increases. However, the capacity is decreased as the power allocated to other beams increases, due to the interbeam interference. As a result, the capacity of each beam is determined not only by the power allocated to it, but also by the power allocated to the other beams.

3. Mathematical Formulation of the Power Allocation

In this paper, the metric to evaluate the power allocation results minimizes the sum of the squared differences between the traffic demand and the capacity allocated to each beam. As a result, the metric will ensure a relatively greater capacity allocation to the beams when there are higher traffic demands, which will achieve greater fairness of the power allocations amongst the beams. At the same time, the metric will also work to maximize the total system capacity. Therefore, the metric considers a compromise between the maximization of total system capacity and the fairness of the power allocations amongst the beams. As a result, the optimization is formulated as follows:

$$\begin{aligned} \min_{\{P_i\}} \quad & \sum_{i=1}^{N} (T_i - C_i)^2 \\ \text{s.t.} \quad & C_i = W \log_2 \left(1 + \frac{P_i}{\gamma W N_0 + \sum_{k=1,k\neq i}^{N} P_k h_{i,k}}\right), \\ & \sum_{i=1}^{N} P_i \leq P_{\text{total}}. \end{aligned} \quad (3)$$

When there is no interbeam interference between the beams, each of the elements in the interbeam interference matrix is equal to zero. As a result, the optimization is convex [2], and the globally optimal solution can be obtained by the optimization. However, when interbeam interference actually exists, it is seen that the optimal variables P_i are coupled with each other. Therefore, it is difficult to determine whether the optimization is convex or not and to obtain the globally optimal power solution for the optimization. To this end, an algorithm based on duality theory is proposed to obtain a locally optimal solution for the optimization [9–12], as presented in the following section.

4. Proposed Dynamic Power Allocation Algorithm

As mentioned above, the proposed dynamic power allocation algorithm is based on duality theory [13]. By introducing the nonnegative dual variable λ, the Lagrange function is given by

$$L(\mathbf{P}, \lambda) = \sum_{i=1}^{N} (T_i - C_i)^2 - \lambda \left(P_{\text{total}} - \sum_{i=1}^{N} P_i\right), \quad (4)$$

where $\mathbf{P} = [P_1, P_2, \ldots, P_N]$.

From (4), the Lagrange dual function can be obtained by

$$D(\lambda) = \min_{\mathbf{P}} L(\mathbf{P}, \lambda), \quad (5)$$

and the dual problem can be written as

$$d^* = \max_{\lambda \geq 0, \sigma_i \geq 0} D(\lambda). \quad (6)$$

The dual problem in (6) can be further decomposed into the following two sequentially iterative subproblems [9].

Subproblem 1: *Power Allocation*. Given the dual variable λ, for any: $i \in \{1, 2, \ldots, N\}$, differentiating (4) with respect to P_i results in the equation below:

$$\frac{\partial D(\boldsymbol{\sigma}, \lambda)}{\partial P_i} = 2 \sum_{j=1}^{N} \left(T_j - C_j\right) \frac{\partial C_j}{\partial P_i} - \lambda = 0. \quad (7)$$

The optimized power allocation of the ith beam P_i can be obtained in (7) by numerical calculation methods, for example, the *golden section*. Moreover, if the optimized P_i is less than zero, then P_i is set to be zero. The detailed expressions in (7) are shown in the appendix.

Subproblem 2: *Dual Variable Update*. The optimal dual variable can be obtained by solving the dual problem:

$$\lambda^{\text{opt}} = \arg \max_{\lambda} \min \left[L\left(\mathbf{P}^{\text{opt}}, \lambda\right)\right]. \quad (8)$$

Because the dual function is always convex, a subgradient method (a generalization of the gradient) can be used here to update the dual variable, as shown below [9]:

$$\lambda^{n+1} = \left[\lambda^n - \Delta_\lambda^n \left(P_{\text{total}} - \sum_{i=1}^{N} P_i^{\text{opt}}\right)\right]^+, \quad (9)$$

where $[x]^+ = \max\{0, x\}$, n is the iteration number, and Δ is the iteration step size.

It has been proven that the above dual variable updating algorithm is guaranteed to converge to the optimal solution, as long as the iteration step chosen is sufficiently small [9].

The whole process of the proposed dynamic power allocation algorithm is summarized as follows.

Step 1. Set appropriate values to λ and P_i, $i \in \{1, 2, \ldots, N\}$.

Step 2. Calculate the value of P_i from (7).

Step 3. Substitute the power values of each beam, as obtained from Step 2, into (9) and then update the dual variable.

Step 4. If the condition of $|\lambda^{n+1}(P_{\text{total}} - \sum_{i=1}^{N} P_i)| < \varepsilon$ is satisfied, then terminate the algorithm; otherwise, jump to Step 2.

Utilizing the above process, the allocated power to each beam is obtained.

5. Simulation Results and Analysis

For the simulation, a multibeam satellite communication system model is set up. The system has 10 beams. For each beam, the bandwidth resource is 50 MHz and the normalized noisepower spectral density parameter γN_0 is $0.2e^{-6}$. Total satellite power is 200 W. The traffic demand of each beam is increased from 80 Mbps to 170 Mbps, by steps of 10 Mbps.

5.1. The Efficiency of the Proposed Power Allocation Algorithm. In order to show the efficiency of the proposed dynamic power allocation algorithm, it is compared with the following two traditional algorithms.

(1) Uniform power allocation algorithm: $P_i = P_{\text{total}}/N$.

(2) Proportional power allocation algorithm: $P_i = P_{\text{total}} T_i / T_{\text{total}}$, where T_{total} is the total traffic demand of all the beams.

Moreover, comparisons are made of the power allocation results for the three algorithms in the following two scenarios, with different interbeam interference matrixes.

Scenario 1. In this system, each beam interferes with the three adjacent beams. As a result, the element in the interbeam interference matrix is set as follows:

$$h_{ij} = \begin{cases} 0.3, & \text{if } |j-i| = 1 \text{ or } |j-i \pm 10| = 1 \\ 0.2, & \text{if } |j-i| = 2 \text{ or } |j-i \pm 10| = 2 \\ 0.1, & \text{if } |j-i| = 3 \text{ or } |j-i \pm 10| = 3 \\ 0, & \text{esle.} \end{cases} \qquad (10)$$

Figure 2 shows the capacity allocated to each beam for the three power allocation algorithms in Scenario 1. Table 1 presents the total system capacities of the three power allocation algorithms in Scenario 1. As shown in Figure 2, the uniform power allocation algorithm uniformly allocates power to each beam, regardless of the traffic demand of the beams or the fairness of the power allocations amongst the beams. Moreover, the total interference from the other beams is the same for each beam, and as a result the capacity allocated to each beam is the same. The proportional power allocation algorithm allocates the power resources to each beam solely according to the traffic demand of each beam, regardless of the interbeam interference. Therefore, the capacity allocated to a beam with high traffic demand is higher than that allocated to a beam with low traffic demand. Compared with the proportional power allocation algorithm, the proposed dynamic power allocation algorithm

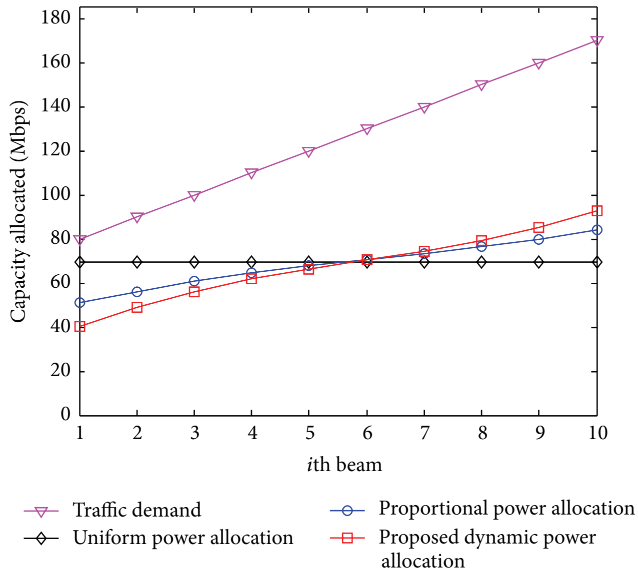


Figure 2: Distribution of the capacity allocated to each beam, for the three algorithms in Scenario 1.

Table 1: Total system capacity of the three algorithms in Scenario 1.

Algorithm	$\sum C_i$
Uniform power allocation	692.83 (Mbps)
Proportional power allocation	685.20 (Mbps)
Proposed dynamic power allocation	675.03 (Mbps)

Table 2: System's total squared difference for the three algorithms from Scenario 1.

Algorithm	$\sum (T_i - C_i)^2$
Uniform power allocation	$3.93E16$
Proportional power allocation	$3.54E16$
Proposed dynamic power allocation	$3.48E16$

allocates more power resources to beams having higher traffic demands, in order to minimize the system's total squared difference between the traffic demand and the capacity allocated to each beam. However, due to the concavity of the capacity function in terms of allocated power, the total system capacity is decreased, which is also shown by the data in Table 1.

Figure 3 shows the squared difference between the traffic demand and allocated capacity of each beam, for the three algorithms from Scenario 1. Table 2 presents the total squared difference for the three algorithms in Scenario 1. As mentioned above, the proposed dynamic power allocation algorithm provides more power resources to the beams with higher traffic demands. As a result, when the results of the proposed dynamic power allocation algorithm are compared to the results of the other two algorithms, the squared difference for the beams with high traffic demands is lower and the squared difference for the beams with low traffic demand is higher. Moreover, the total squared difference for the proposed dynamic power algorithm is the lowest of the

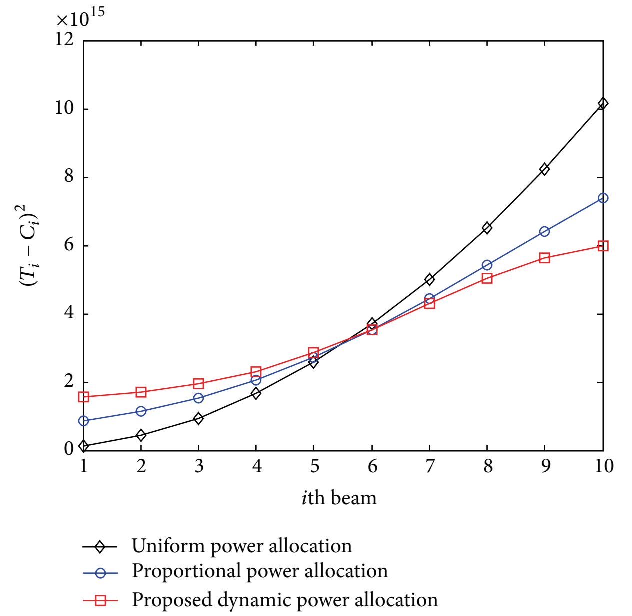


Figure 3: Distribution of the squared difference between the traffic demand and the allocated capacity of each beam, for each of the three algorithms from Scenario 1.

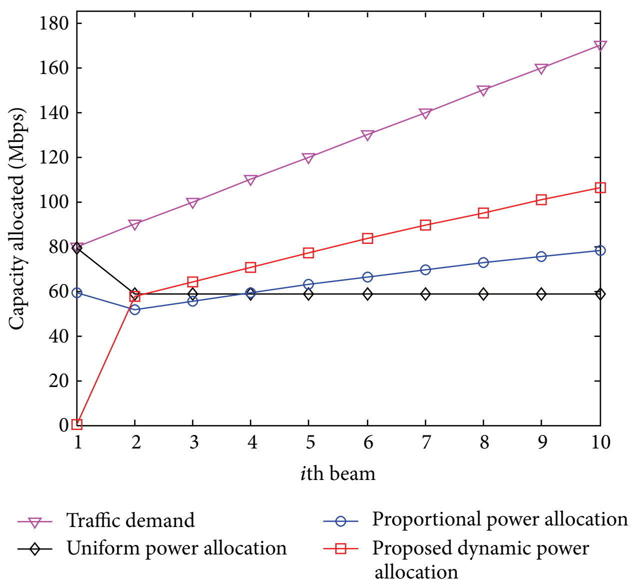


Figure 4: Distribution of the allocated capacity to each beam, for the three algorithms in Scenario 2.

three algorithms. This conclusion is also demonstrated by the data shown in Table 2.

Scenario 2. In this scenario, it is assumed that there is a hostile interference source in Beam 1. Thus we consider that only Beam 1 interferes with the other beams. The interbeam interference matrix is set as follows:

$$h_{ij} = \begin{cases} 0.3, & i = \{2, \ldots 10\}, \ j = 1 \\ 0, & \text{else.} \end{cases} \tag{11}$$

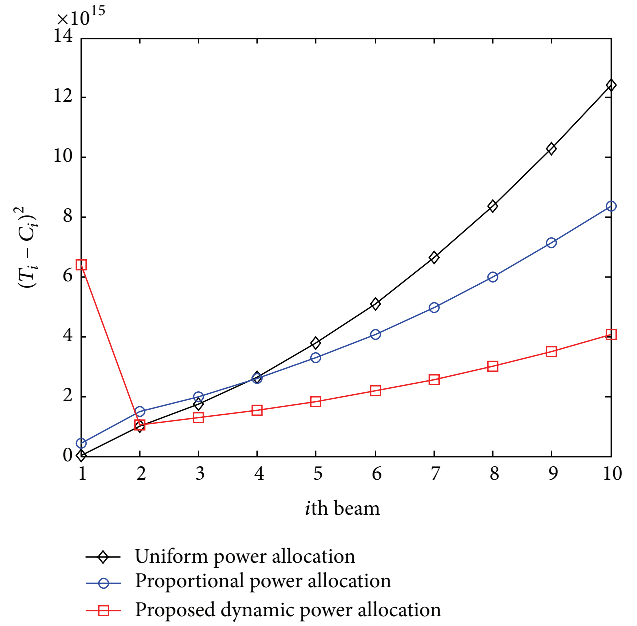


Figure 5: Distribution of the squared difference between the traffic demand and the capacity allocated to each beam, for the three algorithms in Scenario 2.

Figure 4 shows the allocated capacity for each beam, for the three power allocation algorithms in Scenario 2. Table 3 presents the total system capacities of the three power allocation algorithms from Scenario 2. It is known that only Beam 1 interferes with the other beams. It will seem reasonable that, by allocating less power to Beam 1, the interference on the other beams will decrease, and the total system capacity will increase. To this end, the proposed dynamic power allocation algorithm allocates no power to Beam 1 to decrease its interference with the other beams. However, both the uniform and the proportional power allocation algorithms allocate power to each beam, regardless of the interbeam interference matrix in the system. Thus, the power allocated to each beam in Scenario 2 is the same as that in Scenario 1, and the power allocated to Beam 1 is not decreased. As a result, the total system capacity obtained by the two algorithms is less than that obtained by the proposed dynamic power allocation algorithm, as shown in Table 3.

Figure 5 shows the squared difference between the traffic demand and the capacity allocated to each beam, for the three algorithms in Scenario 2. Table 4 presents the total squared difference for the three algorithms in Scenario 2. Figure 5 shows that the squared difference of Beam 1 obtained by the proposed algorithm is higher than that obtained by the other two algorithms. However, the squared differences from Beams 2 to 10 are lower. This is because when compared with the other two algorithms, the proposed dynamic power allocation algorithm provides no power resources to Beam 1 and provides more power to Beams 2 through 10. Moreover, the total squared difference of the proposed dynamic power allocation algorithm is less than that of the other two power allocation algorithms. Taken together with the conclusion

Table 3: Total system capacity for the three algorithms in Scenario 2.

Algorithm	$\sum C_i$
Uniform power allocation	605.71 (Mbps)
Proportional power allocation	650.39 (Mbps)
Proposed dynamic power allocation	744.40 (Mbps)

Table 4: System's total squared difference for the three algorithms in Scenario 2.

Algorithm	$\sum (T_i - C_i)^2$
Uniform power allocation	$5.20E16$
Proportional power allocation	$3.54E16$
Proposed dynamic power allocation	$2.75E16$

Table 5: Total system capacity for the three algorithms in Scenario 3.

Algorithm	$\sum C_i$
Uniform power allocation	519.29 (Mbps)
Proportional power allocation	560.92 (Mbps)
Proposed dynamic power allocation	655.70 (Mbps)

about the total system capacity, it is clear that the proposed dynamic power allocation algorithm improves both the system capacity and the fairness of the power allocations amongst the beams in this scenario.

It is noted that the traffic demand and the channel conditions of each beam are the same in the two scenarios, and only the interbeam interference matrix is different. However, the power allocation result obtained by the proposed algorithm shows a great difference in the two scenarios. In other words, the interbeam interference between the beams has a significant impact on the power allocation results. In addition, the proposed algorithm dynamically allocates the power resource to each beam, taking into account the impact of the interbeam interference between the beams, making the best effort in removing the adverse impacts of the interbeam interference.

5.2. The Effects of the Channel Condition of Each Beam on the Power Allocation Results. It is known that signal attenuation factor γ is affected by channel conditions. To show the impact of the channel conditions of each beam on the power allocation results, the following scenario is set up.

Scenario 3. The normalized noise power spectral density parameters γN_0 from Beams 3 through 5 are set to be $0.2e^{-6}$, $1.2e^{-6}$, and $2.2e^{-6}$. The traffic demand of the three beams is set to be the same as 100 Mbps. The interbeam interference matrix is set to be the same as that in Scenario 2, and other parameters in the system remained the same.

Figure 6 shows the capacity allocated to each beam for the three power allocation algorithms when the channel conditions of each beam are different. Table 5 presents the total system capacity for the three power allocation algorithms in Scenario 3. It is noted that the traffic demand and total interference from the other beams, for Beams 3 through 5, are the same, and only the channel conditions of the three beams are different. As shown in Figure 6, the proposed dynamic power allocation algorithm provides more power resources to the beams that have better channel conditions, and rarely or never provides power to beams with worse channel conditions. Beam 5, for example, with the worst channel condition is provided with no power resources. Therefore, the proposed dynamic power allocation algorithm not only considers the fairness of the power allocations amongst the beams, but also tries to maximize the throughput of the system and achieves a good system performance as predicted. The proportional or uniform power allocation algorithms cannot dynamically allocate the power resources to each beam according to their channel conditions; thus their total system capacities are less than that of the proposed dynamic power allocation algorithm, as clearly demonstrated by the data shown in Table 5.

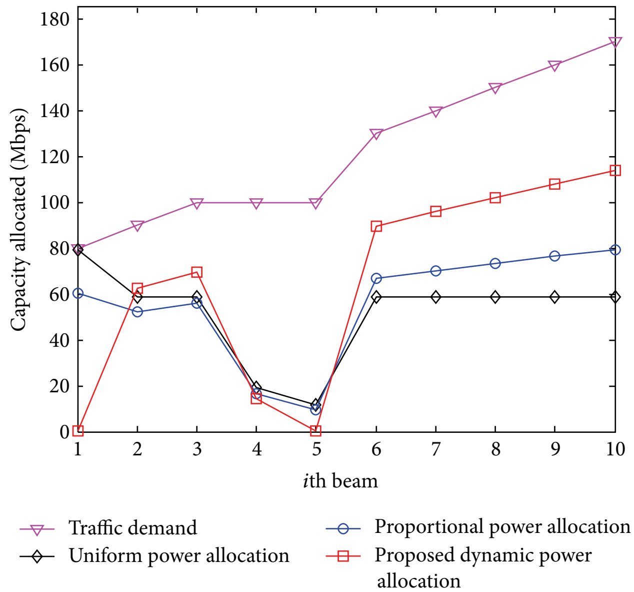


Figure 6: Distribution of the capacity allocated to each beam, for the three algorithms in Scenario 3.

Table 6: System's total squared difference for the three algorithms in Scenario 3.

Algorithm	$\sum (T_i - C_i)^2$
Uniform power allocation	$5.99E16$
Proportional power allocation	$4.88E16$
Proposed dynamic power allocation	$3.71E16$

Figure 7 shows the squared difference between the traffic demand and the capacity allocated to each beam, for the three algorithms in Scenario 3. Table 6 presents the total squared difference of the three algorithms in Scenario 3. As mentioned above, the proposed dynamic power allocation algorithm provides more power resources to the beams having better channel conditions. As a result, the squared difference of the

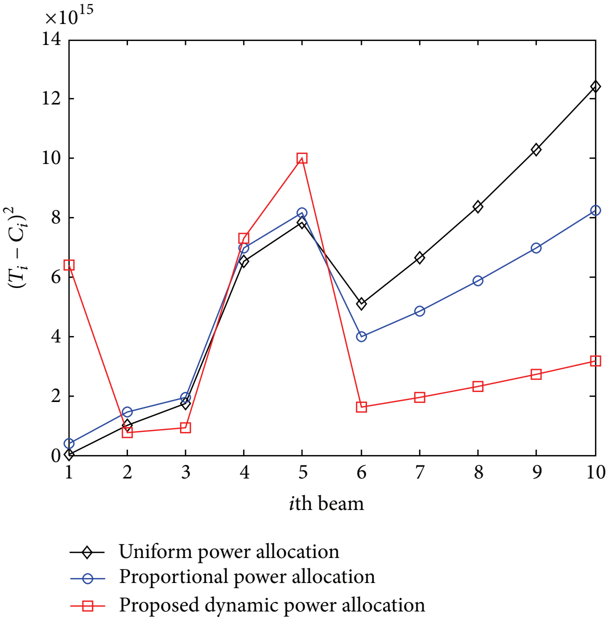


FIGURE 7: Distribution of squared difference between the traffic demand and the capacities allocated to each beam, for the three algorithms in Scenario 3.

traffic demand and the capacity allocated to the beams with better channel conditions is smaller than that of the other two algorithms, as shown in Figure 7. Moreover, the total squared difference of the proposed dynamic power allocation algorithm is also the smallest of the three algorithms. Thus, the fairness of the power allocations amongst the beams that is obtained by the proposed dynamic power allocation algorithm provides the optimal optimization.

6. Conclusions

In multibeam satellite communication systems, due to the reusing of frequencies, there exists interbeam interference between the beams, which cannot be ignored in determining power allocations. To precisely describe the impact of the interbeam interference, the problem of power allocation as a non-linear optimization with constraints was formulated, including a compromise between the maximization of total system capacity and the fairness of the power allocations to the beams. A dynamic power allocation algorithm was then proposed to obtain a locally optimal solution to the optimization.

It was shown that, compared with the traditional uniform or proportional power allocation algorithms, the proposed dynamic power allocation algorithm improved the fairness of the power allocations to the beams and also increased the total system capacity in certain scenarios, such as Scenarios 2 and 3 as presented in Section 5. In addition, the interbeam interference between both the beams and the channel conditions of each beam had a significant impact on the power allocation results. The proposed dynamic power allocation algorithm functioned to remove the adverse impacts of these factors; for example, the algorithm allocated less power to the beams which had greater interference on the other beams, or which had worse channel conditions.

Appendix

Equation (7) is expressed in detail.

When $i = j$, $\partial C_j / \partial P_i$ is given as

$$\frac{\partial C_i}{\partial P_i} = \frac{W}{\ln 2} \cdot \frac{1}{\gamma W N_0 + \sum_{k=1,k\neq i}^{N} P_k h_{ik} + P_i}. \tag{A.1}$$

When $i \neq j$, $\partial C_j / \partial P_i$ is expressed as

$$\begin{aligned}\frac{\partial C_j}{\partial P_i} &= \frac{-W}{\ln 2} \\ &\cdot P_j h_{ji} \times \left(\left(\gamma W N_0 + \sum_{k=1,k\neq j}^{N} P_k h_{jk} \right)^2 \right. \\ &\left. + P_j \left(\gamma W N_0 + \sum_{k=1,k\neq j}^{N} P_k h_{jk} \right) \right)^{-1}.\end{aligned} \tag{A.2}$$

Substituting (A.1) and (A.2) into (7), the following equation is obtained:

$$\begin{aligned}&(T_i - C_i) \frac{2W}{\ln 2} \cdot \frac{1}{\gamma W N_0 + \sum_{k=1,k\neq i}^{N} P_k h_{ik} + P_i} - \lambda \\ &= \sum_{j=1,j\neq i}^{N} \frac{W}{\ln 2} \left(2T_j - 2C_j \right) \cdot M_{ji},\end{aligned} \tag{A.3}$$

where

$$\begin{aligned}M_{ji} = P_j h_{ji} &\times \left(\left(\gamma W N_0 + \sum_{k=1,k\neq j}^{N} P_k h_{jk} \right)^2 \right. \\ &\left. + P_j \left(\gamma W N_0 + \sum_{k=1,k\neq j}^{N} P_k h_{jk} \right) \right)^{-1}.\end{aligned} \tag{A.4}$$

According to (A.3), the optimized power allocation of the ith beam P_i could be obtained by numerical calculation methods.

Conflict of Interests

The authors declare that they do not have any commercial or associative interests that represents a conflict of interest in connection with the work submitted.

Acknowledgment

The authors would like to thank the project support provided by the National High-Tech Research & Development Program of China under Grant 2012AA01A508.

References

[1] G. Maral, M. Bousquet, and Z. Sun, *Communications Systems: Systems, Techniques and Technology*, John Wiley and Sons, New York, NY, USA, 5th edition, 2009.

[2] J. P. Choi and V. W. S. Chan, "Optimum power and beam allocation based on traffic demands and channel conditions over satellite downlinks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2983–2992, 2005.

[3] Y. Hong, A. Srinivasan, B. Cheng, L. Hartman, and P. Andreadis, "Optimal power allocation for multiple beam satellite systems," in *Proceedings of the IEEE Radio and Wireless Symposium (RWS '08)*, pp. 823–826, January 2008.

[4] Q. Feng, G. Li, S. Feng, and Q. Gao, "Optimum power allocation based on traffic demand for multi-beam satellite communication systems," in *Proceedings of the 13th International Conference on Communication Technology (ICCT '11)*, pp. 873–876, September 2011.

[5] U. Park, H. W. Kim, D. S. Oh, and B. J. Ku, "Optimum selective beam allocation scheme for satellite network with multi-spot beams," in *Proceedings of the 4th International Conference on Advances in Satellite and Space Communications (SPACOMM '12)*, pp. 78–81, 2012.

[6] K. Nakahira, K. Kobayashi, and M. Ueba, "Capacity and quality enhancement using an adaptive resource allocation for multi-beam mobile satellite communication systems," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '06)*, pp. 153–158, April 2006.

[7] J. Lei and M. A. Vázquez-Castro, "Joint power and carrier allocation for the multibeam satellite downlink with individual SINR constraints," in *Proceedings of the IEEE International Conference on Communications (ICC '10)*, May 2010.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[9] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, 2006.

[10] R. Wang, V. K. N. Lau, L. Lv, and B. Chen, "Joint cross-layer scheduling and spectrum sensing for OFDMA cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2410–2416, 2009.

[11] G. Ding, Q. Wu, and J. Wang, "Sensing confidence level-based joint spectrum and power allocation in cognitive radio networks," *Wireless Personal Communications*, vol. 72, no. 1, pp. 283–298, 2013.

[12] A. G. Marques, X. Wang, and G. B. Giannakis, "Dynamic resource management for cognitive radios using limited-rate feedback," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3651–3666, 2009.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.