An Optimization Framework for Balancing Throughput and Fairness in Wireless Networks With QoS Support

Ho Ting Cheng, *Student Member, IEEE*, and Weihua Zhuang, *Senior Member, IEEE*

***Abstract*— Quality-of-service (QoS) provisioning, high system throughput, and fairness assurance are indispensable for heterogeneous traffic in future wireless broadband networks. With limited radio resources, increasing system throughput and maintaining fairness are conflicting performance metrics, leading to a natural tradeoff between these two measures. Balancing system throughput and fairness is desired. In this paper, we consider an interference-limited wireless network, and derive a generic optimization framework to obtain an optimal relationship of system throughput and fairness with QoS support and efficient resource utilization, by introducing the bargaining floor. From the relationship curve, different degrees of performance tradeoff between throughput and fairness can be obtained by choosing different bargaining floors. In addition, our framework facilitates call admission control to effectively guarantee QoS of multimedia traffic. The solutions of resource allocation obtained from the optimization framework achieve the Pareto Optimality, demonstrating efficient use of network resources.**

***Index Terms*— Fairness, optimization, Pareto optimality, quality-of-service (QoS), throughput.**

I. INTRODUCTION

FUTURE wireless broadband networks are expected to support ubiquitous communications and mobile computing, the notion of which has been attracting a plethora of attention from academia and industry. Ubiquitous wireless access can be realized in various practical scenarios, namely home networking, office networking, and city networking [1]. Compared with the traditional networking dependent on cables, a wireless network is faster in deployment with lower cost, facilitated by avoiding time-consuming and costly operations such as land construction and cable placement. The wireless networking paradigm provides not only a viable, but also economical solution for both peer-to-peer applications and Internet access. In order to fully exploit promising wireless technologies, efficient and effective radio resource management in wireless networks is crucial.

In wireless networks, system throughput is usually a common performance metric [2]. However, next-generation wireless networks such as wireless mesh networks are anticipated to support multimedia traffic (e.g., voice, video, and data traffic). With heterogeneous traffic, quality-of-service (QoS) provisioning and fairness support are also important. With limited available radio resources, increasing system throughput and maintaining fairness are usually conflicting with each other [3], leading to a natural tradeoff between these two performance measures. In particular, balancing system throughput and fairness with QoS support and high resource efficiency is necessary, depending on different application-specific scenarios [1].

In literature, only limited work addresses the optimal relationship of throughput and fairness [4]–[13]. To the best of our knowledge, there is no widely accepted unified framework to effectively attain different degrees of performance tradeoff between throughput and fairness with QoS support and efficient resource utilization, which is the motivation of this research. With a focus on interference-limited wireless networks, the key contribution of this paper is to derive a unified optimization framework to obtain the optimal relationship (i.e., tradeoff curve) of system throughput and fairness with QoS support. By introducing a bargaining floor, the relationship curve is obtained by solving the optimization problem iteratively. Different degrees of performance tradeoff between throughput and fairness can be achieved by simply choosing appropriate values of the bargaining floor. Given an application of interest, the desired operating tradeoff can be determined. Resource utilization is efficient, which is verified by game theory, achieving Pareto Optimality [14].

To effectively provision QoS of multimedia traffic, *call admission control* is vital to guarantee the QoS requirements of admitted calls. Whether a new incoming call should be admitted or rejected is contingent upon the way how the resources are allocated to wireless links and how much resources are left in the network. Contributing to the criterion of call admission, our optimization framework plays an important role in call admission control so as to facilitate the operation of balancing throughput and fairness.

The rest of the paper is organized as follows. Related work is discussed in Section II. The system model is given in Section III. The optimization problem formulation is presented in Section IV. Efficiency of the proposed resource allocation is addressed in Section V. Numerical results are given in Section VI. Practical implementation issues are discussed in Section VII. Finally, conclusion is drawn in Section VIII.

Manuscript received July 25, 2006; revised December 5, 2006, February 23, 2007, and May 1, 2007; accepted July 5, 2007. The associate editor coordinating the review of this paper and approving it for publication was M. Zorzi. This work was supported in part by the Natural Science and Engineering Research Council (NSERC) of Canada.

The authors are with the Centre for Wireless Communications, Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1 (e-mail: {htcheng,wzhuang}@bbcr.uwaterloo.ca).

Digital Object Identifier 10.1109/TWC.2008.060507.

II. Related Work

In the literature, utility optimization is a tool to measure system performance subject to certain constraints (e.g., QoS requirements) [4]–[7], where a utility function is described as a measure of user satisfaction. Proportional fairness can be obtained by choosing suitable utility functions (e.g., logarithmic functions) [4]. Other performance measure (such as throughput) can also be incorporated into this optimization formulation [6]. With different problem formulations (i.e., utility functions), different optimal solutions can be obtained (e.g., optimal throughput). Pricing schemes [4] can be employed to achieve a tradeoff between throughput and fairness, to a certain extent. However, the utility functions used in these work may not have any physical meaning. How to find a meaningful utility function with an appropriate pricing scheme can also be problematic. In addition, most of the current work assumes that the utility functions can be separable in the dual problem [4]–[7], which may not always be the case, especially for interference-limited systems such as code-division multiple access (CDMA) and ultra-wideband (UWB) systems, meaning that applying existing approaches (e.g., [7]) generally results in suboptimal solutions. In the interference-limited systems, optimality should be obtained by considering all (user) utility functions together.

Ideal (weighted max-min) fairness can be obtained by generalized processor sharing (GPS) [8], [9] or its variants [10], where all nodes in the network share the total resources. With GPS, the resource allocated to each node is dependent on its own weight, whereby each node can have a fair share of resources. However, the notion of a weight is an abstract concept and the question of how to relate QoS requirements to the weight effectively remains unsolved. In GPS, even though all the weights or QoS requirements are already satisfied, only (weighted max-min) fairness is considered. Further, the throughput performance of GPS needs further investigation.

Although some research aims at the tradeoff between throughput and fairness in telecommunications networks, only heuristic schemes are proposed without any optimality consideration [11]–[13]. Thus far, only limited relationship of system throughput and fairness is addressed. Joint consideration of both performance metrics and hence a unified framework attaining different degrees of performance tradeoff between them with QoS support are desired. Further, with scarce radio resources, efficient resource utilization is vital, which can be verified by game theory [14]. All these aspects are addressed in this paper.

III. System Model

We consider a generic system model which is an interference-limited wireless network. We assume that the channel gains are known in advance or can be estimated accurately via pilot symbols. Let R_m^d denote the *effective bandwidth* of a call on the m^{th} link, which is the minimum rate required to satisfy the QoS requirements and depends on source traffic characteristics [16], [17]. Let $R_m(\mathbf{a})$ denote the actual transmission rate of the m^{th} link where $\mathbf{a} = (a_1, a_2, ..., a_m, ..., a_M)$ and a_m is the power scaling factor of the m^{th} link's transmitter, i.e., $a_m \in [0, 1]$, and M the total number of active links in the network. For simplicity, the actual transmission rate of the m^{th} link is given by [18]

$$R_m(\mathbf{a}) = B \log_2 (1 + \gamma_m) \quad (1)$$

where B is the channel bandwidth and γ_m is the signal-to-interference-plus-noise ratio of the m^{th} link. In (1),

$$\gamma_m = \frac{G_{mm} P_m a_m}{\sigma \sum_{n \neq m} G_{mn} P_n a_n + \eta} \quad (2)$$

where P_m is a maximum transmit power level of the m^{th} link's transmitter, G_{mn} is the channel gain from the n^{th} link's transmitter to the m^{th} link's receiver, σ is the cross-correlation factor between any two signals, i.e., $\sigma \in (0, 1]$, and η is the background noise power. Notice that (1) can be easily extended to incorporate bit-error-rate (BER) requirements, coding and modulation schemes to compute the attainable channel transmission rate (or capacity) [19]. This attainable channel transmission rate is equal to the achievable throughput at the physical layer. In practice, the throughput (or goodput) obtained at the medium access control layer (and higher layers) is usually lower than the channel transmission rate due to the overhead of packetization, medium access control, etc. Given all the details of system parameters, the actual throughput can be computed from the channel transmission rate.

In order to effectively balance system throughput and fairness, call admission control is indispensable, which can be contingent on our optimization framework (to be discussed in Section IV). The call admission routine[1] is invoked whenever a new call arrives. For each incoming new call, it is admitted as long as there exists a feasible solution of the optimization problem, meaning that the QoS requirements (i.e., in terms of effective bandwidth) of this new call and all other calls in service can be satisfied. If there is no feasible solution of the optimization problem, the new call is rejected due to insufficient resources available to meet its QoS requirements. In other words, the criterion of call admission is tantamount to the feasibility of the solution of our optimization framework. With the call admission control in place, the QoS requirements of all admitted calls can be guaranteed and the operation of balancing throughput and fairness can be carried out effectively.

Given that the QoS requirements can be met, the network performance can be further improved for increasing system throughput and/or maintaining (weighted max-min) fairness, utilizing the resources efficiently. In this research, the notion of weighted max-min fairness is taken as the fairness performance of interest, which corresponds to the ideal fairness achieved by the GPS. A summary of important symbols used in this paper is given in Table I.

IV. Optimization Problem Formulation

In this section, we first consider two optimization problem formulations, namely system throughput optimization and

[1]Notice that routing is involved in the call admission control. Particularly, the tasks of *QoS routing* include 1) route discovery; 2) call admission control over each link; and 3) route repair. Therefore, which link a call is to traverse along can be determined [15].

TABLE I
SUMMARY OF IMPORTANT SYMBOLS

Symbol	Definition
R_m^d	effective bandwidth of a call on the m^{th} link
$R_m(\mathbf{a})$	actual transmission rate of the m^{th} link
$Q_m(\mathbf{a})$	utility function (i.e., extra throughput obtained) of the m^{th} link, i.e., $Q_m(\mathbf{a}) = R_m(\mathbf{a}) - R_m^d$
w_m	weighting factor of the m^{th} link, i.e., $w_m > 0$
P_m	maximum transmit power level of the m^{th} link's transmitter
G_{mn}	channel gain from the n^{th} link's transmitter to the m^{th} link's receiver
a_m	power scaling factor of the m^{th} link's transmitter, i.e., $a_m \in [0, 1]$
γ_m	signal-to-interference-plus-noise ratio of the m^{th} link
σ	cross-correlation factor between any two signals, i.e., $\sigma \in (0, 1]$
η	background noise power
M	total number of active links in the network
B	channel bandwidth
J	bargaining floor, i.e., $J \in [0, J^*]$, where J^* is the maximum value of J
D	deviation of $\min_m \{w_m Q_m(\mathbf{a})\}$ from J^*, i.e., $D = \lvert J^* - \min_m \{w_m Q_m(\mathbf{a})\}\rvert$
U	measure of system throughput, i.e., $U \in [0, 1]$
V	measure of fairness, i.e., $V \in [0, 1]$
W	shaped Jain's fairness index, i.e., $W \in [0, 1]$

weighted max-min fairness optimization. Next, the generic optimization problem formulation with system throughput and fairness consideration is presented. For the system throughput optimization, the optimization problem is given by

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} R_m(\mathbf{a}) \right\} \tag{3}$$

$$\text{subject to } R_m(\mathbf{a}) \geq R_m^d, 0 \leq a_m \leq 1, \forall m \tag{4}$$

where $\mathbf{a} = (a_1, a_2, ..., a_m, ..., a_M)$ is the *optimization variable*. In fact, (3) can be rewritten as

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} \left(R_m(\mathbf{a}) - R_m^d \right) \right\} \tag{5}$$

as the optimality still maintains for a linear-shifted objective function [20]. The system throughput optimization problem (STOP) can be rewritten as

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \tag{6}$$

$$\text{subject to } Q_m(\mathbf{a}) \geq 0, 0 \leq a_m \leq 1, \forall m \tag{7}$$

where $Q_m(\mathbf{a}) = R_m(\mathbf{a}) - R_m^d$. The physical meaning of $Q_m(\mathbf{a})$ is the amount of extra resources allocated to (i.e., excess throughput obtained by) the m^{th} link. As a comparison, $Q_m(\mathbf{a})$ can be viewed as the utility function of user m in the conventional utility maximization [4]–[6], as $Q_m(\mathbf{a})$ of user m is: 1) increasing; 2) strictly concave; and 3) twice differentiable (i.e., continuous) on a_m. Then, the interpretation of (6) is to achieve social optimality (i.e., maximum of total utility functions). However, in our case, the utility functions are not separable in the dual problem as $Q_m(\mathbf{a})$ only increases with a_m but not over $\mathbf{a}$. Consider the following (partial) dual problem [5]:

$$\min_{\boldsymbol{\lambda} \succeq \mathbf{0}} D(\boldsymbol{\lambda}) \tag{8}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_m, ..., \lambda_M)$ is a set of Lagrange multipliers and

$$D(\boldsymbol{\lambda}) = \max_{\mathbf{a}} \left\{ \sum_m Q_m(\mathbf{a}) + \sum_m \lambda_m Q_m(\mathbf{a}) \right\}. \tag{9}$$

Suppose that the optimality still maintains when each (user) utility function, $Q_m(\mathbf{a})$, is optimized separately. Intuitively, each user m will simply choose the maximum power scaling factor, i.e., $a_m = 1, \forall m$. In most cases, due to the cross-interference (i.e., $\sigma \neq 0$), the overall result will not contribute to the maximum value of (6), though Nash Equilibrium is achieved [14]. Optimizing individual user utility function separately may not always result in the maximal solution of the STOP, as in general from (9),

$$D(\boldsymbol{\lambda}) = \max_{\mathbf{a}} \left\{ \sum_m Q_m(\mathbf{a}) + \sum_m \lambda_m Q_m(\mathbf{a}) \right\} \tag{10}$$

$$= \max_{\mathbf{a}} \left\{ \sum_m (1 + \lambda_m) Q_m(\mathbf{a}) \right\} \tag{11}$$

$$\neq \sum_m \max_{a_m} \{(1 + \lambda_m) Q_m(\mathbf{a})\}. \tag{12}$$

Thus, the solution space (i.e., resource allocation) is not necessarily the same as those proposed in the literature (e.g., [5]). Obtaining the optimal solution requires a joint consideration of all user utility functions, meaning that existing solutions cannot be directly applied.

For the weighted max-min fairness optimization problem

(WMMFOP), the corresponding formulation is given by [21]

$$\max_{\mathbf{a}} \left\{ \min_{m} \{ w_m Q_m(\mathbf{a}) \} \right\} \tag{13}$$

$$\text{subject to } Q_m(\mathbf{a}) \geq 0, 0 \leq a_m \leq 1, \forall m \tag{14}$$

where w_m is a weighting factor of the m^{th} link, i.e., $w_m > 0$, which indicates the *unwillingness* of obtaining extra allocated resources, i.e., the smaller the value of w_m, the more eager is the m^{th} link to obtain more extra resources. The use of w_m is necessary for effective and efficient resource allocation in wireless networks with heterogeneous traffic (e.g., voice, video, and data traffic). For example, for voice traffic, after its effective bandwidth requirement is satisfied, allocating extra resources to it may be wasteful as the quality of signal reception is already good, and hence this traffic should be assigned a larger weighting factor. On the other hand, a smaller weighting factor should be assigned to data traffic, as it demands more throughput even though its effective bandwidth requirement is already met. With different weighting factors, different traffic classes can be differentiated. Notice that the meaning of the weighting factor w_m of our interest is different from that in GPS.

Proposition 1: The set of feasible weighted utilities (i.e., $w_m Q_m(\mathbf{a}), \forall m$) in the WMMFOP has the solidarity property [21].

Proof: Suppose that the feasible resource allocation solution is $\mathbf{a}^*$. Without loss of generality, we assume that there exists an n value such that $Q_n(\mathbf{a}^*) > 0$ (i.e., $a_n^* > 0$), where $n \neq m$ [21]. For a particular timeslot t, over which the m^{th} and n^{th} links are to be active, given the resource allocation solution $\mathbf{a}^*$, the values of their weighted utilities are to be $w_m Q_m(\mathbf{a}^*)$ and $w_n Q_n(\mathbf{a}^*)$, respectively. Let l_t denote the length of the timeslot t, i.e., $l_t > 0$. Similar to [21], it is possible to partition the slot into three minislots, namely t_1, t_2, and t_3 with positive durations l_{t_1}, l_{t_2}, and l_{t_3}, respectively, such that $l_{t_1} + l_{t_2} + l_{t_3} = l_t$. Note that the choice of how to determine these values is arbitrary. During t_1, the resource allocation $\mathbf{a}$ is chosen to be the same as that in timeslot t. During t_2, the same allocation solution is kept as in t_1, except $a_n = 0$. During t_3, we set $a_n = 0$, and a_m is adjusted until the interference experienced by other active links is larger than that in the original timeslot t, if possible, otherwise, we set $a_m = 1$.

In this new resource allocation, compared with the weighted utilities obtained in the original resource allocation in timeslot t, all the links excluding the m^{th} and n^{th} links have the same or higher weighted utilities in t_1 and t_2, respectively. In t_3, their weighted utilities can be higher, the same, or lower, depending on the value of a_m. Since the partitioning into t_1, t_2, and t_3 is entirely arbitrary, it is possible to choose their lengths l_{t_1}, l_{t_2}, and l_{t_3} so that there exist small ϵ_m (> 0) and ϵ_n (> 0) such that the weighted utility of the m^{th} link increases by at most ϵ_m and the weighted utility of the n^{th} link decreases by at most ϵ_n, while the rest of the active links have the same or higher weighted utilities.

Let $w_m Q_m(\tilde{\mathbf{a}})$ and $w_n Q_n(\tilde{\mathbf{a}})$ denote the newly obtained weighted utilities of the m^{th} link and the n^{th} link, respectively. We can now acquire the following inequalities: $w_m Q_m(\mathbf{a}^*) < w_m Q_m(\tilde{\mathbf{a}}) < w_m Q_m(\mathbf{a}^*) + \epsilon_m$ and $w_n Q_n(\mathbf{a}^*) - \epsilon_n < w_n Q_n(\tilde{\mathbf{a}}) < w_n Q_n(\mathbf{a}^*)$. Therefore, the value of $w_m Q_m(\mathbf{a})$ can be increased by at most ϵ_m by decreasing the value of $w_n Q_n(\mathbf{a})$ by at most ϵ_n. And,

$$\begin{bmatrix} \vdots \\ w_m Q_m(\tilde{\mathbf{a}}) \\ w_n Q_n(\tilde{\mathbf{a}}) \\ \vdots \\ w_k Q_k(\tilde{\mathbf{a}}) \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ w_m Q_m(\mathbf{a}^*) \\ w_n Q_n(\mathbf{a}^*) \\ \vdots \\ w_k Q_k(\mathbf{a}^*) \\ \vdots \end{bmatrix} - \begin{bmatrix} \vdots \\ 0 \\ \epsilon_n \\ \vdots \\ 0 \\ \vdots \end{bmatrix} + \begin{bmatrix} \vdots \\ \epsilon_m \\ 0 \\ \vdots \\ 0 \\ \vdots \end{bmatrix} + \sum_{k \neq m,n} \begin{bmatrix} \vdots \\ 0 \\ 0 \\ \vdots \\ \epsilon_k \\ \vdots \end{bmatrix} \tag{15}$$

where $\epsilon_k > 0, \forall k \neq m, n$. Since all $w_m Q_m(\tilde{\mathbf{a}})$, $\forall_m$, belong to the feasible set (i.e., $Q_m(\tilde{\mathbf{a}}) \geq 0$, $\forall_m$), by the definition of solidarity [21], the set of feasible weighted utilities in the WMMFOP has the solidarity property. ■

In fact, the WMMFOP can be transformed into [20]

$$\max_{\mathbf{a}} L \tag{16}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq L, \forall m. \tag{17}$$

Therefore, the WMMFOP can be rewritten as follows:

$$\max_{\mathbf{a}} L \tag{18}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq L, 0 \leq a_m \leq 1, \forall m. \tag{19}$$

With the solidarity property, at the optimal point, all $w_m Q_m(\mathbf{a})$, $\forall m$, are to be equal, and the maximum value of L is unique in WMMFOP [21]. Let J^* and $\hat{\mathbf{a}}$ denote the optimal solutions (i.e., maximal L and optimal $\mathbf{a}$) of the WMMFOP. If the constraints $w_m Q_m(\mathbf{a}) \geq J^*, \forall m$, are added to the STOP, the *modified* STOP becomes

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \tag{20}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq J^*, 0 \leq a_m \leq 1, \forall m. \tag{21}$$

Proposition 2: The optimal solution $\hat{\mathbf{a}}$ obtained from the WMMFOP is also the optimal solution for the modified STOP.

Proof: Suppose that there exists another solution $\tilde{\mathbf{a}}$ such that $\sum_{m=1}^{M} Q_m(\tilde{\mathbf{a}}) > \sum_{m=1}^{M} Q_m(\hat{\mathbf{a}})$. It means that there exists some m such that $w_m Q_m(\tilde{\mathbf{a}}) > J^*$ within the feasible region. In the WMMFOP, if for some m, $w_m Q_m(\tilde{\mathbf{a}}) > J^*$, J^* can be increased by decreasing the value of $\tilde{a}_m$ or increasing the value of $\tilde{a}_n$ or both, for $n \neq m$, within the feasible region until it reaches the maximal value, say $\tilde{J}$. However, it contradicts to the statement that J^* is the maximal value obtained from the WMMFOP. Therefore, no such a solution $\tilde{\mathbf{a}}$ exists. The optimal solution $\hat{\mathbf{a}}$ obtained from the WMMFOP is also the optimal solution for the modified STOP. ■

With the new constraint set, the solution $\mathbf{a}$ obtained from the modified STOP not only achieves weighted max-min fairness, but also optimizes the system throughput. Therefore, to bridge the system throughput and fairness performance measures together, we introduce a parameter called *bargaining floor*, denoted by J, where $J \in [0, J^*]$ and J^* is solution (i.e., the maximal value) of the WMMFOP. Motivated by Proposition 2, we propose the generic optimization problem (GOP), which is given by

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \tag{22}$$

$$\text{subject to } w_m Q_m(\mathbf{a}) \geq J, 0 \leq a_m \leq 1, \forall m. \tag{23}$$

Clearly, the solutions of the GOP for maximal system throughput are obtained when $J = 0$ while that for maximal (weighted max-min) fairness when $J = J^*$, where J^* is obtained from the WMMFOP. In this research, our focus is not to solve the GOP. Instead, we employ it as a unified framework for deducing the optimal relationship between system throughput and fairness with QoS support. Let $\mathbf{a}^*$ be the optimal solution obtained from the GOP.

Proposition 3: The system throughput (i.e., $\sum_{m=1}^{M} Q_m(\mathbf{a}^*)$) is a non-increasing function of bargaining floor J.

Proof: When J increases (decreases), the feasible region of $\mathbf{a}$ in the GOP shrinks (expands). For $0 \leq J_1 \leq J_2 \leq J^*$, let $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$ denote the optimal solutions of the GOP with J_1 and J_2, respectively. The feasible region of $\mathbf{a}$ of the GOP with J_2 is only a subset of that with J_1. Thus, $\sum_{m=1}^{M} Q_m(\mathbf{a}_1^*) \geq \sum_{m=1}^{M} Q_m(\mathbf{a}_2^*)$ and hence the system throughput does not increase with the value of J. ■

Corollary 1: The minimum value of $w_m Q_m(\mathbf{a}^*)$ (i.e., $\min_m \{w_m Q_m(\mathbf{a}^*)\}$) is a non-decreasing function of J.

Proof: Let J^* be the solution of WMMFOP and $\mathbf{a}_1^*$ be the optimal solution of the GOP with J_1, where $0 \leq J_1 \leq J^*$. For $0 \leq J_1 < J_2 \leq J^*$, consider the following two cases:

Case 1: If the solution $\mathbf{a}_1^*$ is feasible for the GOP with J_2, from Proposition 3, $\mathbf{a}_1^*$ is also the optimal solution for the GOP with J_2. Hence, the minimum value of $w_m Q_m(\mathbf{a}^*)$ is the same for the GOP with J_1 and J_2.

Case 2: If the solution $\mathbf{a}_1^*$ is not feasible for the GOP with J_2, it means that there exists some m such that $w_m Q_m(\mathbf{a}_1^*) < J_2$ and hence $\min_m \{w_m Q_m(\mathbf{a}_1^*)\} < J_2$. Thus, $\mathbf{a}_1^*$ is an infeasible solution for the GOP with J_2. In addition, suppose that $\mathbf{a}_2^*$ is the optimal solution for the GOP with J_2, i.e., $w_m Q_m(\mathbf{a}_2^*) \geq J_2, \forall m$. Hence, $\min_m \{w_m Q_m(\mathbf{a}_2^*)\} \geq J_2 > \min_m \{w_m Q_m(\mathbf{a}_1^*)\}$.

By combining the above two cases, $\min_m \{w_m Q_m(\mathbf{a}_2^*)\} \geq \min_m \{w_m Q_m(\mathbf{a}_1^*)\}$ and hence the minimum value of $w_m Q_m(\mathbf{a}^*)$ (i.e., $\min_m \{w_m Q_m(\mathbf{a}^*)\}$) does not decrease with the value of J. ■

Theorem 1: The system throughput (i.e., $\sum_{m=1}^{M} Q_m(\mathbf{a}^*)$) does not increase with J, but the minimum value of $w_m Q_m(\mathbf{a}^*)$ (i.e., $\min_m \{w_m Q_m(\mathbf{a}^*)\}$) does not decrease with J.

Proof: By Proposition 3 and Corollary 1, it is proved. ■

Corollary 2: A relationship between the system throughput and weighted max-min fairness performance can be achieved by solving the GOP with different values of J.

Proof: From Theorem 1, for $J = 0$ the solution obtained from the GOP corresponds to the maximal system throughput, while for $J = J^*$ the solution obtained from the GOP corresponds to the maximal weighted max-min fairness performance. When J increases from zero, the solution obtained from the GOP can refer to decreased system throughput and increased fairness performance. Therefore, the performance tradeoff and hence a desired relationship between the system throughput and weighted max-min fairness performance can be achieved by solving the GOP with different values of J, i.e., $J \in [0, J^*]$. ■

From the perspective of network design, with a limited amount of resources, improving fairness performance will reduce the system throughput, which matches with the perspective of Corollary 2. With different values of J, the tradeoff curve of system throughput and fairness can be obtained. Thus, the GOP should be solved with different values of J iteratively. The iterative procedure to obtain the tradeoff curve is described below:

Step 1: Find J^* by solving the WMMFOP;

Step 2: Set $J = 0$ and solve the GOP, whereby the obtained solution $\mathbf{a}$ corresponds to the maximal throughput performance;

Step 3: Increase J by δJ and solve the GOP again;

Step 4: Repeat Step 3 until $J = J^*$, which corresponds to the maximal fairness.

Through the above procedure, different sets of the optimal solution $\mathbf{a}$, the corresponding relationships of system throughput and fairness, and hence the desired tradeoff can be obtained by suitably selecting the value of J. Notice that after the tradeoff curve is procured, the choice of the J value usually depends on the purpose of the application of interest and/or the prerogative of the system designer. With the fixed value of J, the existence of the optimal solution of the GOP (i.e., the desired tradeoff) is assured due to the call admission control in place, discussed in Section III. In case that the value of J is allowed to change randomly when the system is in use, a conservative approach can be employed: whenever a new call arrives, we only check the feasibility condition on the solution of the WMMFOP (i.e., the GOP with $J = J^*$). The new call is admitted if there exists a solution, or rejected otherwise. Since the feasible region does not shrink and hence the solution obtained initially will not become infeasible when the value of J decreases, this prevents the ongoing transmissions from being dropped. In this case, the same call admission control is employed, guaranteeing the QoS requirements of all calls in service with any value of J.

V. Efficiency Evaluation by Game Theory

In this section, we show that our solutions obtained from the GOP given by (22)–(23) achieve efficient use of resources. In game theory, efficient resource utilization is determined by the concept of Pareto Optimality [14].

Definition 1: An action profile $\mathbf{b}^* = (b_1^*, b_2^*, ..., b_M^*)$ is said to be *Pareto optimal* if and only if there exists no other action profile $\tilde{\mathbf{b}}$ such that for some m, $Y_m(\tilde{\mathbf{b}}) > Y_m(\mathbf{b}^*)$ and $Y_n(\tilde{\mathbf{b}}) \geq Y_n(\mathbf{b}^*)$, for $n \neq m$, where $Y_m(\cdot)$ is a payoff of user m in the context of game theory. In words, an action profile (or resource allocation) is Pareto optimal if there exists no other action profile that makes some user(s) better off without making the other user(s) worse off.

Proposition 4: The optimal solution $\mathbf{a}$ of the GOP is Pareto optimal.

Proof: Given the action profile or optimal solution $\mathbf{a}^*$ (i.e., resource allocation) obtained from the GOP, denote $Q_m(\mathbf{a}^*)$ or $w_m Q_m(\mathbf{a}^*)$ as the payoff of the m^{th} link. Consider another action profile $\tilde{\mathbf{a}}$. For some m, if $Q_m(\tilde{\mathbf{a}}) > Q_m(\mathbf{a}^*)$, then $Q_n(\tilde{\mathbf{a}}) < Q_n(\mathbf{a}^*)$ for some n, as either $\tilde{a}_m > a_m^*$ or $\tilde{a}_n < a_n^*$ or both. According to Definition 1, the optimal solution $\mathbf{a}^*$ obtained from the GOP achieves the Pareto Optimality. ■

From the perspective of game theory, the resources are efficiently utilized for increasing system throughput and/or maintaining fairness. In other words, for the relationship of system throughput and fairness, every point (i.e., resource allocation) on the tradeoff curve (discussed in Section VI) is Pareto optimal, utilizing the resources efficiently.

VI. Numerical Results

This section presents numerical results on: 1) system throughput and fairness performance versus the value of J in the GOP; and 2) the desired relationship (i.e., tradeoff curve) of system throughput and fairness. In the numerical analysis, we simply solve the GOP by an exhaustive search with an increment size of $\delta a = 0.01$. Suppose that there are I iterations in the iterative procedure for the throughput and fairness tradeoff curve. Let $\mathbf{a_i}$ be the optimal solution obtained from the GOP in the i^{th} iteration. The measure of system throughput in the i^{th} iteration (i.e., $i \in I$) is given by

$$U = \frac{\left(\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right) - \min_{i\in I}\left\{\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right\}}{\max_{i\in I}\left\{\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right\} - \min_{i\in I}\left\{\sum_{m=1}^{M} Q_m(\mathbf{a_i})\right\}} \tag{24}$$

where $U \in [0,1]$, i.e., the larger the value of the system throughput, the larger the value of U. Let $D_i = |J^* - \min_m\{w_m Q_m(\mathbf{a_i})\}|$, where J^* is the solution of the WMMFOP and D_i represents the deviation of the minimum value of $w_m Q_m(\mathbf{a_i})$ among all M links from J^*, i.e., the larger the value of D_i, the poorer the weighted max-min fairness performance. The measure of weighted max-min fairness in the i^{th} iteration is given by

$$V = \frac{\max_{i\in I}\{D_i\} - D_i}{\max_{i\in I}\{D_i\} - \min_{i\in I}\{D_i\}} \tag{25}$$

where $V \in [0,1]$, i.e., the larger the value of D_i, the smaller the value of V. V indicates the fairness performance of the worst link. In the literature, Jain's fairness index [22] is widely employed as a measure of network-wise fairness performance. Let JFI_i be the Jain's fairness index in the i^{th} iteration, where

$$JFI_i = \frac{\left(\sum_m w_m Q_m(\mathbf{a_i})\right)^2}{M\sum_m \left(w_m Q_m(\mathbf{a_i})\right)^2}. \tag{26}$$

TABLE II
System Parameters for the Numerical analysis for the Case of Equal Weighting Factors

Parameter	Value
M	4
B	1
η	0.01
σ	0.1
P_m	1, $\forall m$
w_m	1, $\forall m$
$(R_1^d, ..., R_4^d)$	(2, 1, 0.5, 0.1)
J^*	0.3290

The shaped Jain's fairness index in the i^{th} iteration is given by

$$W = \frac{JFI_i - \min_{i\in I}\{JFI_i\}}{\max_{i\in I}\{JFI_i\} - \min_{i\in I}\{JFI_i\}} \tag{27}$$

where $W \in [0,1]$, i.e., the larger the value of Jain's fairness index, the larger the value of W.

In this numerical analysis, we consider four active links in the network, i.e., $M = 4$. The fading coefficient of a link is modeled as a complex Gaussian random variable with zero mean and unit variance. The channel gain matrix $\mathbf{G}$, i.e., $\mathbf{G} = [G_{mn}]_{M\times M}$, used for the numerical analysis is randomly generated and normalized, which is given below:

$$\mathbf{G} = \begin{bmatrix} 0.2818 & 0.3299 & 0.2739 & 0.0350 \\ 0.2418 & 0.1761 & 0.5019 & 1.0000 \\ 0.1823 & 0.9345 & 0.2802 & 0.0068 \\ 0.2016 & 0.4150 & 0.4480 & 0.0400 \end{bmatrix}. \tag{28}$$

Under the condition that the solution exists in the WMMFOP, we consider two cases: 1) all the weighting factors are equal; and 2) the weighting factors are different.

A. Equal Weighting Factors

In this case, the system parameters are given in Table II, where J^* is computed by solving the WMMFOP. We follow the iterative procedure described in Section IV and obtain the numerical results given in Table III.

First, we study the behaviors of the system throughput measure U and the fairness measure V with different values of J, which is given in Fig. 1. As mentioned in Section IV, the tradeoff curve is obtained by solving the GOP iteratively, starting at the maximal system throughput and ending at the maximal fairness. As expected, in Fig. 1, U decreases from the maximum value (i.e., $U = 1$) to the minimum value (i.e., $U = 0$) with J, while V increases from the minimum value (i.e., $V = 0$) to the maximum value (i.e., $V = 1$) with J, which shows that increasing system throughput and maintaining fairness are conflicting with each other. The shaped Jain's fairness index is also plotted for comparison.

From the results shown in Table III, the minimum weighted utility value, $\min_m\{w_m Q_m(\mathbf{a})\}$, increases with J, as expected. Note that the fairness performance measure of V and that of W are different (i.e., V for the worst-link fairness performance while W for the network-wise fairness performance); however, in the case of equal weighting factors, the

TABLE III

NUMERICAL RESULTS FOR THE CASE OF EQUAL WEIGHTING FACTORS

J	$Q_1(\mathbf{a})$	$Q_2(\mathbf{a})$	$Q_3(\mathbf{a})$	$Q_4(\mathbf{a})$	$\sum Q_m(\mathbf{a})$	$\min_m \{w_m Q_m(\mathbf{a})\}$	U	V	W	$\mathbf{a} = [a_1, a_2, a_3, a_4]$
0	0.9115	0.0063	1.3134	0.0001	2.2313	0.0001	1.0000	0	0	[1.00, 0.46, 0.64, 0.14]
0.0219	0.8548	0.0230	1.2921	0.0236	2.1935	0.0230	0.9584	0.0696	0.0272	[1.00, 0.50, 0.66, 0.18]
0.0439	0.8716	0.0478	1.1874	0.0457	2.1525	0.0457	0.9131	0.1385	0.0892	[1.00, 0.52, 0.61, 0.21]
0.0658	0.8809	0.0674	1.0971	0.0672	2.1126	0.0672	0.8693	0.2041	0.1445	[1.00, 0.54, 0.57, 0.24]
0.0877	0.7930	0.0885	1.0959	0.0900	2.0674	0.0885	0.8194	0.2686	0.1792	[1.00, 0.60, 0.61, 0.29]
0.1097	0.8088	0.1097	0.9938	0.1117	2.0240	0.1097	0.7717	0.3331	0.2488	[1.00, 0.62, 0.56, 0.32]
0.1316	0.8997	0.1353	0.8077	0.1321	1.9748	0.1321	0.7175	0.4012	0.3198	[1.00, 0.61, 0.45, 0.33]
0.1535	0.8662	0.1553	0.7520	0.1541	1.9276	0.1541	0.6656	0.4682	0.3786	[1.00, 0.65, 0.44, 0.37]
0.1755	0.7390	0.1775	0.7835	0.1766	1.8766	0.1766	0.6093	0.5367	0.4544	[1.00, 0.74, 0.50, 0.44]
0.1974	0.8167	0.1986	0.6100	0.1983	1.8236	0.1983	0.5509	0.6025	0.5004	[1.00, 0.73, 0.40, 0.45]
0.2193	0.7079	0.2227	0.6189	0.2195	1.7690	0.2195	0.4908	0.6671	0.6038	[1.00, 0.82, 0.44, 0.52]
0.2413	0.6777	0.2417	0.5501	0.2425	1.7120	0.2417	0.4281	0.7345	0.6755	[1.00, 0.87, 0.42, 0.57]
0.2632	0.6693	0.2646	0.4535	0.2643	1.6517	0.2643	0.3616	0.8032	0.7262	[1.00, 0.91, 0.38, 0.61]
0.2851	0.6609	0.2864	0.3569	0.2858	1.5900	0.2858	0.2936	0.8685	0.7433	[1.00, 0.95, 0.34, 0.65]
0.3071	0.5515	0.3081	0.3187	0.3098	1.4881	0.3081	0.1815	0.9365	0.8596	[0.94, 1.00, 0.33, 0.70]
0.3290	0.3312	0.3292	0.3340	0.3290	1.3234	0.3290	0	1.0000	1.0000	[0.78, 1.00, 0.33, 0.71]

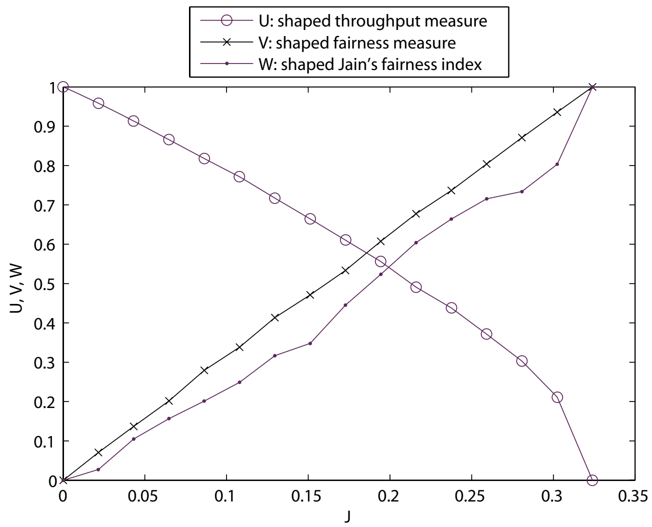

Fig. 1. The system throughput measure and the fairness measures against the value of J with equal weighting factors.

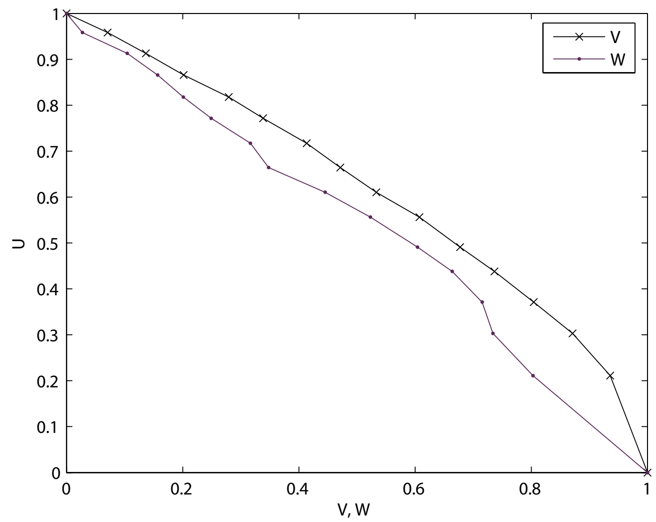

Fig. 2. The relationship of system throughput and fairness with equal weighting factors.

general trend of both curves agrees with each other. Thus, both fairness measures match with the max-min fairness performance, as both V and W increase with $\min_m \{w_m Q_m(\mathbf{a})\}$, in general.

Consider the trend of each utility function. For every link, its utility value (i.e., $Q_m(\mathbf{a})$ of the m^{th} link) against the value of J is given in Table III. For a small J, the links with smaller effective bandwidths usually obtain larger utility values (e.g., $Q_3(\mathbf{a})$). It is intuitive that those links with smaller effective bandwidths have more freedom to increase their throughputs than other links with larger effective bandwidths. As J increases, the utility values of those links with smaller effective bandwidths decrease for the sake of achieving a certain level of fairness. However, with different channel gains, some link, say the m^{th} link, with a small effective bandwidth may be forced to use a small value of a_m so that only a small value of $Q_m(\mathbf{a})$ is achieved, for example, $Q_4(\mathbf{a})$ in our example. From (28), $G_{24} = 1.0$, meaning that the interference impact from the 4^{th} link to the 2^{nd} link is significant. In order to meet all the effective bandwidth requirements, the 4^{th} link can only use a small value of a_4, which results in a small value of $Q_4(\mathbf{a})$. Nonetheless, the utility values of all links converge to the same value when the condition of maximal fairness is met (i.e., weighted max-min fairness) as all the weighting factors are equal. Notice that the discrepancies in Table III are merely due to the discrete exhaustive search used in the numerical analysis.

The desired tradeoff curve of system throughput and fairness performances is shown in Fig. 2. The curve is a bit concave in shape, meaning that in a nearly unfair situation (i.e., $V \approx 0$), a unit decrease in system throughput gives a larger marginal improvement in weighted max-min fairness performance. At a near-maximal fairness point (i.e., $V \approx 1$), a larger decrease of system throughput is required to further increase the fairness measure. From this curve, different de-

TABLE IV
SYSTEM PARAMETERS FOR THE NUMERICAL ANALYSIS FOR THE CASE OF DIFFERENT WEIGHTING FACTORS

Parameter	Value
M	4
B	1
η	0.01
σ	0.1
P_m	1, $\forall m$
$(w_1, ..., w_4)$	(1, 2, 4, 8)
$\left(R_1^d, ..., R_4^d\right)$	(2, 1, 0.5, 0.1)
J^*	0.9249

grees of performance tradeoff between system throughput and fairness can be found by suitably choosing the value of J. The shaped Jain's fairness index is also plotted for reference. This tradeoff curve is undoubtedly useful for effective and efficient resource allocation. With application-specific constraints (such as fairness or throughput requirements), a desired tradeoff point can be obtained from this relationship curve and hence the corresponding resource allocation $\mathbf{a}$ can be deduced.

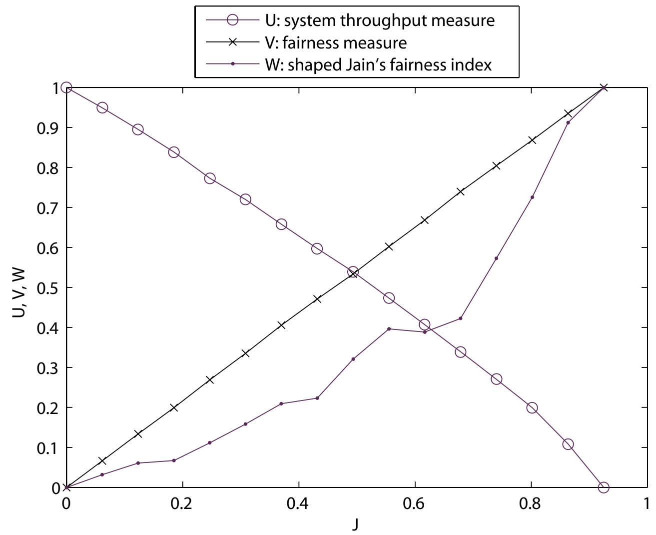

Fig. 3. The system throughput measure and the fairness measures against the value of J with different weighting factors.

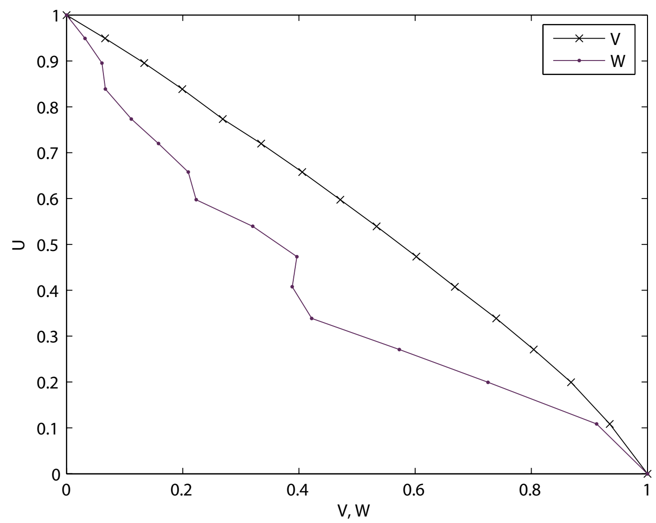

Fig. 4. The relationship of system throughput and fairness with different weighting factors.

B. Different Weighting Factors

For the case of different weighting factors, the details of the system parameters for the numerical analysis are given in Table IV and the numerical results are given in Table V.

In Fig. 3, the curve of the system throughput measure U and that of the fairness measure V with different values of J are plotted, and the tradeoff curve of system throughput and fairness performances is shown in Fig. 4. For U and V, the trends of these two curves in Fig. 3 and Fig. 4 are more or less the same as those shown in Fig. 1 and Fig. 2, respectively. For W, however, the curve of W deviates more away from that of V in the case of different weighting factors. In fact, using W in this case is less effective to accurately indicate the improvement of weighted max-min fairness performance though the minimum weighted utility value, $\min_m \{w_m Q_m(\mathbf{a})\}$, increases with J, shown in Table V. Nonetheless, the Jain's fairness index is a network-wise fairness measure, instead of measuring the worst-link fairness performance, though the general trend of the curve agrees with the weighted max-min fairness performance. Notice that the fluctuations in the graph are partly because of the discrete exhaustive search used in the numerical analysis.

With different weighting factors, the allocation solution $\mathbf{a}$ in the GOP is not the same as that with the equal weighting factor. As mentioned, the smaller the value of w_m, the more eager is the m^{th} link to obtain more extra resources. At the last row of Table V, the value of $Q_1(\mathbf{a})$ is the largest while that of $Q_4(\mathbf{a})$ is the smallest, as $w_1 < w_2 < w_3 < w_4$. In fact, with different weighting factors, different traffic classes can be differentiated. Therefore, the use of w_m is crucial for further improving the effectiveness and efficiency of resource allocation in wireless networks with heterogeneous traffic (e.g., voice, video, and data traffic).

VII. DISCUSSIONS ON PRACTICAL IMPLEMENTATION

In this section, some issues on practical implementation are discussed. For wireless networks with centralized control, a central controller (such as a base station) essentially collects requests from wireless nodes, makes decisions on call admission, and allocates resources to wireless links. The global network information (i.e., network resources, QoS requirements of calls, and channel conditions) are available at the central controller, thereby leading to an easier practical implementation to obtain (near-)optimal resource allocation by solving the GOP. In contrast, for wireless networks with distributed control (without central coordination), acquiring global network information by message exchanges is not desired, causing a considerable amount of overhead. Thus, each node usually has its local network information only. Node clustering is a viable approach, where the whole network is divided into clusters. Within each cluster, a clusterhead merely

TABLE V
NUMERICAL RESULTS FOR THE CASE OF DIFFERENT WEIGHTING FACTORS

J	$Q_1(\mathbf{a})$	$Q_2(\mathbf{a})$	$Q_3(\mathbf{a})$	$Q_4(\mathbf{a})$	$\sum Q_m(\mathbf{a})$	$\min_m\{w_m Q_m(\mathbf{a})\}$	U	V	W	$\mathbf{a}=[a_1,a_2,a_3,a_4]$
0	0.9115	0.0063	1.3134	0.0001	2.2313	0.0010	1.0000	0	0	[1.00, 0.46, 0.64, 0.14]
0.0617	0.9328	0.0313	1.2341	0.0089	2.2071	0.0625	0.9492	0.0666	0.0318	[1.00, 0.47, 0.60, 0.15]
0.1233	0.9285	0.0623	1.1746	0.0159	2.1813	0.1246	0.8952	0.1338	0.0610	[1.00, 0.49, 0.58, 0.16]
0.1850	0.8412	0.0941	1.1959	0.0232	2.1544	0.1853	0.8385	0.1994	0.0670	[1.00, 0.54, 0.63, 0.18]
0.2466	0.8522	0.1317	1.1082	0.0312	2.1233	0.2496	0.7733	0.2690	0.1117	[1.00, 0.56, 0.59, 0.19]
0.3083	0.8724	0.1553	1.0300	0.0401	2.0978	0.3106	0.7198	0.3351	0.1585	[1.00, 0.57, 0.55, 0.20]
0.3700	0.8759	0.1880	0.9567	0.0477	2.0683	0.3759	0.6579	0.4058	0.2096	[1.00, 0.59, 0.52, 0.21]
0.4316	0.8069	0.2199	0.9581	0.0546	2.0395	0.4364	0.5974	0.4713	0.2231	[1.00, 0.64, 0.55, 0.23]
0.4933	0.8754	0.2471	0.8272	0.0621	2.0118	0.4941	0.5394	0.5337	0.3208	[1.00, 0.63, 0.47, 0.23]
0.5550	0.8789	0.2786	0.7529	0.0698	1.9802	0.5573	0.4733	0.6021	0.3965	[1.00, 0.65, 0.44, 0.24]
0.6166	0.7774	0.3093	0.7837	0.0785	1.9489	0.6186	0.4076	0.6685	0.3887	[1.00, 0.72, 0.49, 0.27]
0.6783	0.7276	0.3426	0.7604	0.0856	1.9162	0.6844	0.3386	0.7397	0.4223	[1.00, 0.77, 0.50, 0.29]
0.7399	0.7910	0.3720	0.6265	0.0942	1.8837	0.7441	0.2708	0.8043	0.5729	[1.00, 0.76, 0.42, 0.29]
0.8016	0.8345	0.4045	0.5103	0.1004	1.8497	0.8035	0.1993	0.8686	0.7255	[1.00, 0.76, 0.36, 0.29]
0.8633	0.8969	0.4324	0.3678	0.1094	1.8065	0.8649	0.1086	0.9350	0.9124	[1.00, 0.75, 0.29, 0.29]
0.9249	0.9260	0.4625	0.2452	0.1211	1.7548	0.9249	0	1.0000	1.0000	[1.00, 0.76, 0.24, 0.30]

deals with the network activity of its own neighborhood [23]. Each clusterhead can run the GOP locally by treating the inter-cluster interference as a portion of the intra-cluster interference, yet leading to a suboptimal solution.

Whether a resource allocation solution obtained is optimal or suboptimal depends on the algorithm design. Although our framework is universal in the sense that it can be applied to any system model which is interference-limited (e.g., CDMA systems), one drawback is that solving this optimization problem generally involves high computational complexity. Some (suboptimal) algorithms with low complexity are preferred for the sake of practical implementation. To tackle a non-trivial optimization problem (i.e., non-convex in nature) such as the GOP, two approaches are commonly used:

1) Convex approximation. The original optimization problem is relaxed to a convex optimization problem. The optimal solution of the relaxed problem can be achieved by classical methods such as gradient methods and interior-point methods [24]. In [7], it is proved that under a high signal-to-interference-plus-noise ratio approximation, the objective function of the GOP can be formulated to be a concave function (by verifying that the Hessian matrix is negative definite) and hence the GOP becomes a convex optimization problem. A rich body of literature exists on the theory of convex optimization whereby fast, simple, and robust practical algorithms can be devised (e.g., gradient-based iterations) [7], [24]; and
2) Interpreting Karush-Kuhn-Tucker (KKT) conditions [24]. The necessary conditions for the optimal solution can be verified the by KKT conditions. The KKT conditions for the GOP are as follows:

$$\sum_{m=1}^{M} \frac{\partial Q_m(\mathbf{a})}{\partial a_m} + \alpha_m w_m \frac{\partial Q_m(\mathbf{a})}{\partial a_m} = \lambda_m - \mu_m, \ \forall m \tag{29}$$

$$w_m Q_m(\mathbf{a}) \geq J, \ \forall m \tag{30}$$

$$\alpha_m \left(J - w_m Q_m(\mathbf{a})\right) = 0, \ \forall m \tag{31}$$

$$\lambda_m \left(a_m - 1\right) = 0, \ \forall m \tag{32}$$

$$\mu_m a_m = 0, \ \forall m \tag{33}$$

$$\alpha_m, \lambda_m, \mu_m \geq 0, \ \forall m \tag{34}$$

$$0 \leq a_m \leq 1, \ \forall m \tag{35}$$

where α_m is the Lagrange multiplier for the constraint (30), and λ_m and μ_m are the Lagrange multipliers for the constraint (35). Optimal and/or heuristic algorithms can be deduced based on the structure of the KKT conditions or conceived as methods for solving the KKT conditions [24], [25]. For example, combining (29) and (33), we have

$$\left\{\lambda_m - (1+\alpha_m w_m)\frac{\partial Q_m(\mathbf{a})}{\partial a_m} - \sum_{n\neq m}^{M} \frac{\partial Q_n(\mathbf{a})}{\partial a_m}\right\} \cdot a_m = 0, \ \forall m \tag{36}$$

and hence

$$\lambda_m \geq (1+\alpha_m w_m)\frac{\partial Q_m(\mathbf{a})}{\partial a_m} - \left\{-\sum_{n\neq m}^{M} \frac{\partial Q_n(\mathbf{a})}{\partial a_m}\right\}, \ \forall m. \tag{37}$$

Therefore, if a_m is positive, equality holds for (37), which is the necessary condition of a_m being positive, providing some guidelines for the design of a resource allocation algorithm. To handle a large amount of decision variables, *data structure* plays an important role in determining the complexity. For instance, tree implementation together with sorting and/or searching can facilitate to bring lower computational complexity to the algorithm [26].

In this paper, the goal is to develop a unified framework to balance throughput and fairness. Nonetheless, the actual system performance with practical implementation needs further investigation. Devising practical algorithms and actual system operations (such as medium access control) tailored for our system with accurate traffic models is left for future work.

VIII. Conclusion

In this paper, we propose the unified optimization framework for interference-limited wireless networks, whereby the optimal relationship curve of system throughput and fairness can be obtained. Different degrees of performance tradeoff between system throughput and fairness can be achieved, by suitably adjusting the value of the bargaining floor. QoS support is assured with the help of the call admission control based on the feasibility of the solution of the GOP. From the perspective of game theory, the resource allocation solutions achieve the Pareto Optimality, demonstrating efficient use of network resources.

Acknowledgement

The authors wish to thank the anonymous reviewers for their helpful reviews and suggestions which improved the quality and presentation of this paper.

References

[1] H. T. Cheng, H. Jiang, and W. Zhuang, "Distributed medium access control for wireless mesh networks," *Wireless Commun. Mobile Computing*, vol. 6, no. 6, pp. 845–864, Sep. 2006.

[2] R. Jurdak, C. V. Lopes, and P. Baldi, "A survey, classification and comparative analysis of medium access control protocols for ad hoc networks," *IEEE Commun. Surveys Tutorials*, vol. 6, no. 1, pp. 2–16, 2004.

[3] H. Luo, J. Cheng, and S. Lu, "Self-coordinating localized fair queueing in wireless ad hoc networks," *IEEE Trans. Mobile Computing*, vol. 3, no. 1, pp. 86–98, Jan.–Mar. 2004.

[4] F. P. Kelly, A. Maulloo, and D. K. H. Tan, "Rate control in communication networks: Shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[5] S. H. Low and D. E. Lapsely, "Optimization flow control I: Basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, no. 6, pp. 861–874, Dec. 1999.

[6] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," in *Proc. IEEE INFOCOM*, June 2002, vol. 2, pp. 477–486.

[7] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Select. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.

[8] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case," *IEEE/ACM Trans. Networking*, vol. 1, no. 3, pp. 344–357, June 1993.

[9] ——, "A generalized processor sharing approach to flow control in integrated services networks: The multiple node case," *IEEE/ACM Trans. Networking*, vol. 2, no. 2, pp. 137–150, 1994.

[10] S. Lu, V. Bharghavan, and R. Srikant, "Fair scheduling in wireless packet networks," *IEEE/ACM Trans. Networking*, vol. 7, no. 4, pp. 473–489, Aug. 1999.

[11] G. Miklos and S. Molna, "Fair allocation of elastic traffic for a wireless base station," in *Proc. IEEE Globecom*, 1999, vol. 3, pp. 1673–1678.

[12] J. Damnjanovic, A. Jain, T. Chen, and S. Sarkar, "Scheduling the CDMA2000 reverse link," in *Proc. IEEE VTC*, 2002, vol. 1, pp. 386–390.

[13] B. Wang, K. C. Chua, and H.-P. Schwefel, "Resource allocation for non-realtime data users in UMTS uplink," in *Proc. IEEE WCNC*, Mar. 2003, vol. 2, pp. 1055–1059.

[14] G. Owen, *Game Theory*, 3rd ed. San Diego: Academic Press, 2001.

[15] A. Abdrabou and W. Zhuang, "A position-based QoS routing scheme for UWB mobile ad hoc networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 4, pp. 850–856, Apr. 2006.

[16] F. P. Kelly, "Notes on effective bandwidths," in *Stochastic Networks: Theory and Applications*. Oxford, U.K.: Oxford University Press, 1996.

[17] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.

[18] C. E. Shannon, "A mathematical theory of communications," *Bell Labs Technical J.*, vol. 27, pp. 623–656, 1948.

[19] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2004.

[20] H. M. Wagner, *Principles of Operations Research*. Upper Saddle River, NJ: Prentice Hall, Inc., 1975.

[21] B. Radunović and J. Y. Le Boudec, "Rate performance objectives of multihop wireless networks," *IEEE Trans. Mobile Computing*, vol. 3, no. 4, pp. 334–349, Oct.–Dec. 2004.

[22] R. Jain, A. Durresi, and G. Babic, "Throughput fairness index: An explanation," ATM Forum Document Number: ATM Forum/99-0045, Feb. 1999.

[23] J. Yu and P. Chong, "A survey of clustering schemes for mobile ad hoc networks," *IEEE Commun. Surveys Tutorials*, vol. 7, no. 1, pp. 32–48, First Qtr. 2005.

[24] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[25] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Select. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[26] M. A. Weiss, *Data Structures and Algorithm Analysis in C*. Menlo Park, CA: Addison Wesley, 1997.

Ho Ting Cheng (S'05) received the BEng and MPhil degrees in Information Engineering from The Chinese University of Hong Kong (CUHK) in 2003 and 2005, respectively. He is a research assistant and currently working toward his PhD degree at the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include distributed resource allocation, wireless mesh networking, and communication theory.

Weihua Zhuang (M'93-SM'01) is a Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. She is a co-author of the textbook *Wireless Communications and Networking* (Prentice Hall, 2003). Her current research interests include wireless communications and networks, and radio positioning.

Dr. Zhuang is a co-recipient of a Best Paper Award from IEEE ICC 2007 and a Best Student Paper Award from IEEE WCNC 2007. She received the Outstanding Performance Award in 2005 and 2006 from the University of Waterloo for outstanding achievements in teaching, research, and service, and the Premier's Research Excellence Award (PREA) in 2001 from the Ontario Government for demonstrated excellence of scientific and academic contributions. She is the Editor-in-Chief of the IEEE Transactions on Vehicular Technology, and an Editor of the IEEE Transactions on Wireless Communications, the *EURASIP Journal on Wireless Communications and Networking*, and the *International Journal of Sensor Networks*.