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

Ho Ting Cheng
Department of Electrical and Computer Engineering, University of Waterloo, Canada
htcheng@bbcr.uwaterloo.ca

Weihua Zhuang
Department of Electrical and Computer Engineering, University of Waterloo, Canada
wzhuang@bbcr.uwaterloo.ca

ABSTRACT

Throughput and fairness are conflicting performance metrics, leading to a natural tradeoff between these two measures. In this paper, we derive a generic optimization framework to obtain a relationship of system throughput and fairness, 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. The solutions of resource allocation obtained from the optimization framework achieve the Pareto Optimality, demonstrating efficient use of network resources.

1. INTRODUCTION

In wireless networking, system throughput is usually the common performance metric for network design [4]. Future wireless broadband networks are expected 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 [5], 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.

In literature, only limited work on the optimal relationship of throughput and fairness is addressed [1, 3, 7]. Utility optimization is a tool to measure system performance subject to certain constraints (e.g., QoS requirements) [1, 3], where a utility function is described as a measure of user satisfaction. With different problem formulations (i.e., utility functions), different performance measures can be obtained (e.g., maximal throughput). Pricing schemes [3] 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 are separable in the dual problem [1, 3], which may not always be the case, especially for interference-limited systems (such as code-division multiple access (CDMA) systems), meaning that applying existing approaches (e.g., [1]) generally results in suboptimal solutions. Ideal (max-min) fairness can be obtained by generalized processor sharing (GPS) [7], 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 (max-min) fairness is considered. Throughput performance is not addressed properly in this context.

To the best of our knowledge, there is no 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. The key contribution of this paper is to derive a unified optimization framework to obtain optimal relationship (i.e., tradeoff curve) of system throughput and fairness with QoS support for interference-limited wireless networks by introducing the bargaining floor. Resource utilization is efficient, which is verified by game theory, achieving the Pareto Optimality [6].

2. SYSTEM MODEL

We consider a generic system model which is interference-limited for wireless networks. Assume that the channel gains are known. We further assume that call admission control is in place so that the QoS requirements (i.e., required transmission rate) of all admitted calls can be met by suitable resource allocation. Let R_m^d be the required transmission rate of m^{th} link, $R_m(\mathbf{a})$ be the actual transmission rate of m^{th} link where $\mathbf{a} = (a_1, a_2, ..., a_m, ..., a_M)$ and a_m is the power scaling factor of m^{th} link's transmitter, i.e., $a_m \in [0, 1]$, and M be the total number of active links in the network. For simplicity, the actual transmission rate of m^{th} link is given by

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

where B is the channel bandwidth and γ_m is the signal-to-

*This research was supported by a research grant from the Bell University Laboratories (BUL).

QShine'06, The Third International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, August 7-9 2006
Waterloo, Canada, © 2006 ACM 1-59593-472-3/06/08...$5.00

interference-plus-noise ratio of 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 m^{th} link's transmitter, G_{mn} is the channel gain from n^{th} link's transmitter to 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 extended to incorporate required bit-error-rate (BER) requirements, coding and modulation schemes easily [1].

After the QoS requirements are met, the network performance can be further improved for increasing system throughput and/or maintaining (max-min) fairness.

3. PROBLEM FORMULATION

In this section, we first consider two optimization problem formulations, namely system throughput optimization and max-min fairness optimization. Next, the generic optimization problem formulation with system throughput and fairness consideration is presented.

The system throughput optimization problem (STOP) is given by

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

$$\text{subject to } R_m(\mathbf{a}) \geq R_m^d, 0 \leq a_m \leq 1, \forall_m. \quad (4)$$

The STOP can be rewritten as

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \quad (5)$$

$$\text{subject to } Q_m(\mathbf{a}) \geq 0, 0 \leq a_m \leq 1, \forall_m \quad (6)$$

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 (i.e., excess throughput obtained) to 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 [3]. 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 $\mathbf{a}$. Thus, the solution space (i.e., resource allocation) is not necessarily the same as those proposed in literature (e.g., [1]), meaning that existing approaches cannot be directly applied.

For the max-min fairness, the corresponding optimization problem formulation is given by [8]

$$\max_{\mathbf{a}} \left\{ \min_{m} Q_m(\mathbf{a}) \right\} \quad (7)$$

$$\text{subject to } Q_m(\mathbf{a}) \geq 0, 0 \leq a_m \leq 1, \forall_m. \quad (8)$$

In fact, the max-min fairness optimization problem (MMFOP) can be transformed into [3]

$$\max L \quad (9)$$

$$\text{subject to } Q_m(\mathbf{a}) \geq L, 0 \leq a_m \leq 1, \forall_m. \quad (10)$$

At the maximum point, all $Q_m(\mathbf{a}), \forall_m$, are to be equal, and the maximum value of L is unique. The proof can be found in [8]. Let J^* and $\mathbf{a}^*$ be the optimal solutions (i.e., maximal L and optimal $\mathbf{a}$) of the MMFOP. If the constraints $Q_m(\mathbf{a}) \geq J^*, \forall_m$, are added to the STOP, the optimal solution $\mathbf{a}^*$ obtained from the MMFOP is also the optimal solution for the STOP.

Proposition 1. *If the constraints $Q_m(\mathbf{a}) \geq J^*, \forall_m$, are added to the STOP, where J^* is the maximal value obtained from the MMFOP, the optimal solution $\mathbf{a}^*$ obtained from the MMFOP is also the optimal solution for the STOP.*

Proof. If the constraints $Q_m(\mathbf{a}) \geq J^*, \forall_m$, are added to the STOP, the optimization formulation becomes:

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \quad (11)$$

$$\text{subject to } Q_m(\mathbf{a}) \geq J^*, 0 \leq a_m \leq 1, \forall_m. \quad (12)$$

Suppose that there exists another solution $\tilde{\mathbf{a}}$ such that $\sum_m Q_m(\tilde{\mathbf{a}}) > \sum_m Q_m(\mathbf{a}^*)$. It means that there exists some m such that $Q_m(\tilde{\mathbf{a}}) > J^*$ within the feasible region. In the MMFOP, if for some m, $Q_m(\tilde{\mathbf{a}}) > J^*$ within the feasible region, J^* can be increased by decreasing the value of $\tilde{a}_m$ or increasing the value of $\tilde{a}_n$ or both, for $n \neq m$, until it reaches the maximal value, say $\tilde{J}$, leading to the contradiction. Therefore, no such a solution $\tilde{\mathbf{a}}$ exists. The optimal solution $\mathbf{a}^*$ obtained from the MMFOP is also the optimal solution for the STOP. □

With the new constraint set, the solution $\mathbf{a}$ obtained from the STOP not only achieves max-min fairness, but also optimizes the corresponding 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 the solution (i.e., the maximal value) of the MMFOP. The generic optimization problem (GOP) is given by

$$\max_{\mathbf{a}} \left\{ \sum_{m=1}^{M} Q_m(\mathbf{a}) \right\} \quad (13)$$

$$\text{subject to } Q_m(\mathbf{a}) \geq J, 0 \leq a_m \leq 1, \forall_m. \quad (14)$$

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

Proposition 2. *The system throughput (i.e., $\sum_{m=1}^{M} Q_m(\mathbf{a})$) decreases with J.*

Proof. Omitted due to the space limit. □

Corollary 1. *The minimum value of $Q_m(\mathbf{a})$ (i.e., $\min_m \{Q_m(\mathbf{a})\}$) increases with J.*

Proof. Omitted due to the space limit. □

From the perspective of network resources, with a limited amount of resources, improving fairness performance will reduce the system throughput, which matches with the perspective of the GOP. 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 procedure to obtain the tradeoff curve, denoted by $(*)$, is described below:

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

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 procedure $(*)$, different sets of the optimal solution $\mathbf{a}$[1] and the corresponding relationships of system throughput and fairness can be obtained. We start at the maximal throughput performance and end at the maximal fairness performance. Interestingly, each link seems to bargain with other links based on some agreement (i.e., the bargaining level) and updates its rate iteratively until all links achieve the maximal (max-min) fairness. Therefore, different degrees of performance tradeoff between system throughput and fairness can be found by suitably selecting the values of J.

4. EFFICIENCY EVALUATION

In game theory, efficient resource utilization is determined by the concept of Pareto Optimality [6].

PROPOSITION 3. *The optimal solution* $\mathbf{a}$ *of the GOP is Pareto optimal.*

PROOF. Omitted due to the space limit. □

From the perspective of game theory, the resources are efficiently utilized for increasing system throughput and/or maintaining fairness. In other words, for the optimal relationship of system throughput and fairness, every point (i.e., resource allocation) on the tradeoff curve (discussed in Section 5) is Pareto optimal, utilizing the resources efficiently.

5. NUMERICAL RESULTS

This section presents numerical results on: 1) system throughput and fairness performance with the value of J in the GOP; and 2) the desired 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 $\delta a = 0.025$. Suppose that there are I iterations in the procedure $(*)$. 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

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

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

$$V = \frac{\max_{i\in I}\{D_i\} - D_i}{\max_{i\in I}\{D_i\} - \min_{i\in I}\{D_i\}} \in [0,1].$$

V indicates the fairness performance of the worst link. In literature, Jain's fairness index [2] 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 $JFI_i = \frac{\left(\sum_m Q_m(\mathbf{a_i})\right)^2}{M\sum_m (Q_m(\mathbf{a_i}))^2}$. The shaped Jain's fairness index in the i^{th} iteration is given by

$$W = \frac{JFI_i - \min_{i\in I}\{JFI_i\}}{\max_{i\in I}\{JFI_i\} - \min_{i\in I}\{JFI_i\}} \in [0,1].$$

In this numerical analysis, we consider four active links (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 $\mathbf{G}$ used for the numerical analysis is shown below:

$$\mathbf{G} = \begin{bmatrix} 0.2818 & 0.3299 & 0.2739 & 0.0350 \\ 0.2418 & 0.1761 & 0.5019 & 1.0000 \\ 0.1823 & 0.9345 & 0.2802 & 0.0068 \\ 0.2016 & 0.4150 & 0.4480 & 0.0400 \end{bmatrix}. \quad (15)$$

Other system parameters are chosen as: $B = 1$, $\eta = 0.01$, $\sigma = 0.1$, $P_m = 1, \forall_m$, $\left(R_1^d, ..., R_4^d\right) = (2, 1, 0.5, 0.1)$, and $J^* = 0.3249$, where J^* is computed by solving the MMFOP. We follow the procedure $(*)$ described in Section 3 and the numerical results are shown in Table 1.

First, we study the behaviors of the system throughput measure U and the fairness measure V with different values of J, shown 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 1, the minimum utility value, $\min_m\{Q_m(\mathbf{a})\}$, increases and hence the max-min fairness performance improves 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, the general trend of both curves agrees with each other, though more fluctuations are observed in the curve of W. Thus, both fairness measures match with the max-min fairness performance, as both V and W increase with $\min_m\{Q_m(\mathbf{a})\}$, in general.

Consider the trend of each utility function (i.e., $Q_m(\mathbf{a})$ of m^{th} link) in Table 1. For a small J, the links with smaller required transmission rates usually obtain larger utility values. It is intuitive that those links with smaller required transmission rates have more freedom to increase their throughputs than other links with higher required transmission rates. As J increases, the utility values of those links with smaller required transmission rates decrease for the sake of achieving a certain level of fairness. However, with different channel gains, some link, say m^{th} link, with a small required transmission rate 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 numerical analysis. From (15), $G_{24} = 1.0$, meaning that the interference impact of the 4^{th} link on the 2^{nd} link is significant. In order to meet all the transmission rate 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 J^* when the condition of maximal fairness is met (i.e., max-min fairness). Note that the discrepancies in Table 1 are 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

[1]If there are multiple solutions $\mathbf{a}$ obtained from the GOP, the one (ones) optimizing both system throughput and fairness performance is (are) chosen.

Table 1: **Numerical Results**

J	$Q_1(\mathbf{a})$	$Q_2(\mathbf{a})$	$Q_3(\mathbf{a})$	$Q_4(\mathbf{a})$	$\min_m \{Q_m(\mathbf{a})\}$	U	V	W	$\mathbf{a} = [a_1, a_2, a_3, a_4]$
0	0.8692	0.0051	1.3483	0.0042	0.0042	1.0000	0	0	[1.0000, 0.4750, 0.6750, 0.1500]
0.0433	0.9389	0.0542	1.1099	0.0452	0.0452	0.9136	0.1279	0.1215	[1.0000, 0.5000, 0.5500, 0.2000]
0.0866	0.8605	0.0936	1.0240	0.0888	0.0888	0.8243	0.2639	0.2120	[1.0000, 0.5750, 0.5500, 0.2750]
0.1300	0.8052	0.1363	0.9054	0.1309	0.1309	0.7264	0.3952	0.3255	[1.0000, 0.6500, 0.5250, 0.3500]
0.1733	0.8708	0.1770	0.6531	0.1753	0.1753	0.6147	0.5337	0.4327	[1.0000, 0.6750, 0.4000, 0.4000]
0.2166	0.6970	0.2201	0.6330	0.2200	0.2200	0.4980	0.6730	0.6079	[1.0000, 0.8250, 0.4500, 0.5250]
0.2599	0.6809	0.2646	0.4501	0.2614	0.2614	0.3737	0.8020	0.7171	[1.0000, 0.9000, 0.3750, 0.6000]
0.3032	0.5984	0.3039	0.3059	0.3078	0.3039	0.2187	0.9345	0.8066	[0.9750, 1.0000, 0.3250, 0.7000]
0.3249	0.3274	0.3395	0.3249	0.3252	0.3249	0	1.0000	1.0000	[0.7750, 1.0000, 0.3250, 0.7000]

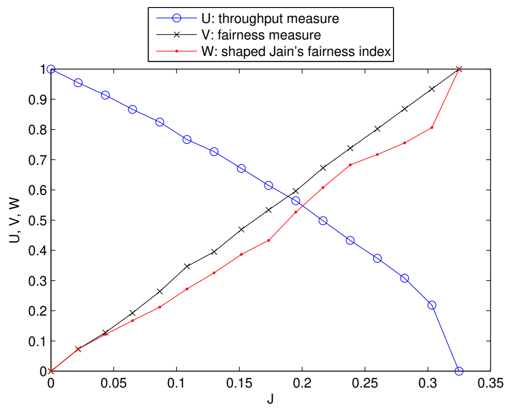

Figure 1: The system throughput measure and the fairness measures against the value of J.

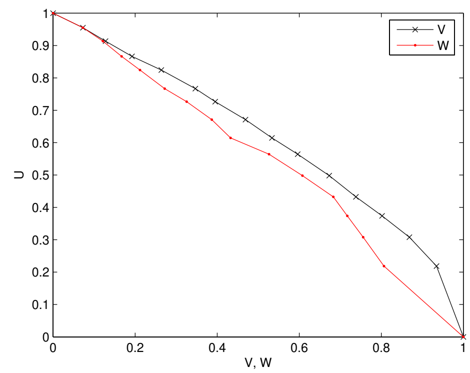

Figure 2: The optimal relationship of system throughput and fairness.

(i.e., $V \approx 0$), a unit decrease in system throughput gives a larger marginal improvement in fairness performance. At a near-maximal fairness point (i.e., $V \approx 1$), a larger decrease in system throughput is required to further increase the fairness measure. From this relationship curve, different degrees of tradeoff between system throughput and fairness performances 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), the optimal tradeoff point can be obtained and hence the corresponding resource allocation **a** can be deduced.

6. 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 with QoS support can be achieved, by suitably adjusting the value of the bargaining floor. From the perspective of game theory, the resource allocation solutions achieve the Pareto Optimality, demonstrating efficient allocation of network resources.

7. REFERENCES

[1] M. Chiang. Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control. *IEEE J. Select. Areas Commun.*, 23(1):104–116, Jan. 2005.

[2] R. Jain, A. Durresi, and G. Babic. Throughput fairness index: an explanation. *ATM Forum Document Number: ATM Forum/990045*, Feb. 1999.

[3] 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 *Proceedings of IEEE INFOCOM '2002*, volume 2, pages 477–486, June 2002.

[4] R. Jurdak, C. V. Lopes, and P. Baldi. A survey, classification and comparative analysis of medium access control protocols for ad hoc networks. *IEEE Communications Surveys and Tutorials*, 6(1), 2004.

[5] H. Luo, J. Cheng, and S. Lu. Self-coordinating localized fair queueing in wireless ad hoc networks. *IEEE Trans. Mobile Comput.*, 3(1):86–98, Jan. 2004.

[6] G. Owen. *Game Theory, 3rd Ed.* Academic Press, 2001.

[7] 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*, 1(3):344–357, June 1993.

[8] B. Radunović and J. Y. Le Boudec. Rate performance objectives of multihop wireless networks. *IEEE Trans. Mobile Comput.*, 3(4):334–349, Oct. 2004.