This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2008 proceedings.

How Bad Is Suboptimal Rate Allocation?

Tian Lan[1], Xiaojun Lin[2], Mung Chiang[1], Ruby Lee[1]
[1]Department of Electrical Engineering, Princeton University, NJ 08544, USA
[2]School of Electrical and Computer Engineering, Purdue University, IN 47907, USA

***Abstract*—Not too bad. A rate allocation that is suboptimal with respect to a utility maximization formulation still maintains the maximum flow-level stability when the utility gap is sufficiently small, and provides a minimum size of stability region otherwise. Utility-suboptimal allocation may also enhance other network performance metrics, e.g., it may increase network throughput and reduce link saturation. Quantifying these intuitions, this paper provides a theoretical support for turning attention from optimal but complex solutions of network optimization to those that are simple even though suboptimal.**

I. Introduction

The framework of Network Utility Maximization (NUM) has been very extensively studied over the last decade since [1]. Formulating many resource allocation problems as maximization of an increasing and concave utility function over a convex constraint set, a large number of publications have developed iterative, distributed algorithms that converge to the optimum. Achieving optimality is clearly nice. Not only it attains the benchmark of the highest value of network utility, it also guarantees flow-level stochastic stability. The number of flows varies over time as they are randomly generated by users and served by the network. This system can be viewed as a queuing network where the service rate depends on the rate allocation policy (or, in general terms, the resource allocation policy) employed by the network. For convex NUM, Markov arrival, and fast-time solution of NUM, it has been shown that for all rate allocation policies maximizing α-fair utilities with $\alpha > 0$, flow-level stochastic stability can be achieved if and only if the traffic intensity lies within the rate region, see, e.g., [2], [3], [4], [5]. In other words, rate region in the α-fair utility maximization problem is also the maximum stability region under arrival and departure dynamics. Utility-optimality and flow-level stability are strong benefits of optimizing NUM.

But in practice it may not be easy to solve NUM optimally. There are many situations where only suboptimal solutions to the utility maximization problem are realistically computable. First, rate allocation algorithms require certain convergence time to compute the optimal rate allocations. If the network configuration (e.g., the number of active users) changes faster than the convergence time, the rate allocations will never be optimal. Second, solving some NUM problems for rate allocation may require solving non-convex scheduling problems even if the resulting feasible rate region is convex. For example, when the feasible rate region of a network is obtained by time-sharing among different subsets of users, a non-convex multi-user scheduling problem still needs to be solved in order to find the exact rate region achieved by time-sharing [6].

Between optimality and simplicity, which one should we pick? Driven by the practical need for simple yet suboptimal solutions, we try to address the question in the title of this paper, by investigating the effects of utility-gap on flow-level stability and on other important network performance metrics such as total throughput and link saturation.

In [7], the authors show that for a class of rate allocation algorithms based on the so-called dual solutions, the optimal stability region can be achieved even if the algorithm does not converge to the optimal rate allocation at any time. Similar observations have also been made in switching [8] and scheduling [9] problems. In this work, we take a different approach. We characterize the capability of a rate allocation algorithm by the gap between its utility and the optimal utility, and we study stability as a function of the utility gap. Thus, our results are not restricted to one type of rate allocation algorithms. Consider a network model with Poisson arrivals and exponential file-size distributions. We first investigate the flow-level stochastic stability for networks with suboptimal rate-allocation policies. Intuitively, one would think that the maximum stability region may be retained if the utility gap is 'small', while only a reduced stability region can be achieved if the utility gap is 'large'. This is indeed true. In Section III, we show that when the ratio of utility gap (caused by a suboptimal rate allocation policy) to the maximum utility approaches zero as queue length tends to infinity, the maximum stability region can be retained. When the utility gap is in proportion to the maximum utility, only a reduced stability region can be achieved. We can still provide a lower bound for the achievable stability region under rate allocation policies satisfying the utility gap condition. These results characterize the stability of a broad class of suboptimal rate allocation policies.

On the other hand, since suboptimal rate allocations with a small enough utility gap is capable of achieving the maximum stability region, we investigate the potential *benefits* of allowing such a utility gap. It is clear that by deliberately *under-optimizing* an α-fair utility, we can achieve network performance improvement in other metrics. However, it is unclear how much improvement we can possibly achieve by under-optimizing the utility with a given allowable gap. We formulate the potential performance improvement as a function of given utility gap, and derive a first-order approximation for these tradeoff curves based on local sensitivity (shadow price) analysis. This formulation generalizes that in [14],

This work has been in part supported by the National Science Foundation through awards CCF-0448012, CNS-0430487, CNS-0752961, CNS-0519880, CCF-0635202, CNS-0720570, and CNS-0721484.

978-1-4244-2026-1/08/$25.00

which considered the tradeoff between total throughput and an α-fairness parameter, and assumed that optimality always holds. Our result not only illustrates the potential benefits of under-optimizing an α-fair utility, but also quantitatively characterizes the *tradeoff* between sacrificing utility value and improving other network performance metrics, especially the total throughput and link saturation.

Proofs of the main results are collected in the Appendix. Vectors are denoted in small letter, e.g., x, with their ith component denoted by x_i. Matrices are denoted by capitalized letters, e.g., A, with A_{ij} denoting the $\{i, j\}$th component. Vector inequalities denoted by $x \succeq y$ are considered componentwise. We use $D(x)$ to denote a diagonal matrix whose diagonal elements are the corresponding components from vector x. Subscripts $(\cdot)^T$ denotes the matrix transpose. We use $\mathcal{R}$ to denote a set of vectors and $\check{\mathcal{R}}$ for its interior.

II. Utility Maximization and Gap

Consider a communication network shared by a set of data flows, which belong to N distinctive flow classes. All flows within the same class have the same resource requirements. Let x_i denote the number of flows of class i that remain in the system. We refer to the vector $x = [x_1, \ldots, x_N]$ as the network state. The problem of network rate allocation is to determine the total rate allocated to class-i flows in state x, denoted by $\phi_i(x)$. Rate $\phi_i(x)$ is equally shared by all class-i flows, each assigned a rate $\phi_i(x)/x_i$. We refer to the vector $\phi(x) = [\phi_1(x), \ldots, \phi_N(x)]$ as the rate allocation in state x. The allocation vector $\phi(x)$ is constrained to lie in a set $\mathcal{R} \subset \mathbb{R}^N_+$.

The set $\mathcal{R}$ represents the physical, technological, and economic constraints of the network under consideration. A rate allocation $\phi(x)$ is feasible if $\phi(x) \in \mathcal{R}$, i.e., there exists a resource allocation and routing policy that can support the network under flow-class rate $\phi(x)$. In this paper, we only require the set $\mathcal{R}$ to be convex, which holds in many settings, e.g., [4], [6].

Various network rate control policies can be derived as solving some utility maximization problem with different utility functions. Given a parameter $\alpha \geq 0$, the α-fair rate allocation [10] is the solution to the following optimization problem

$$\begin{array}{ll} \text{maximize} & \sum_i x_i U_i(\frac{\phi_i}{x_i}) \\ \text{subject to} & \phi \in \mathcal{R} \\ \text{variables} & \phi \end{array} \quad (1)$$

where

$$U_i(y) = \begin{cases} \frac{y^{1-\alpha}}{1-\alpha}, & \alpha \neq 1 \\ \log y, & \alpha = 1 \end{cases} \quad (2)$$

Parameter $\alpha \geq 0$ is a fixed constant modulating the level of fairness, which includes several special cases such as proportional fairness and max-min fairness.

Let $\phi_{\mathrm{opt}}(x)$ denote the optimal solution for the rate allocation problem (1) at state x. It has also been shown in [2], [3], [5] that such rate allocation achieves the maximum stability region (i.e., the interior of the feasible rate region). However, in practical networks, due to the issues of convergence time and complexity of computing the optimal resource allocation, such as schedules, we accept the possibility that only suboptimal solutions to the utility maximization problem (1) may be practically computable, thus the resulting suboptimal rate allocation could possibly reduce network stability and performance. For an arbitrary suboptimal solution $\phi(x)$, we consider a utility gap defined by

$$\Delta(x) = \sum_i x_i U_i(\frac{\phi_{\mathrm{opt},i}}{x_i}) - \sum_i x_i U_i(\frac{\phi_i}{x_i}) \quad (3)$$

The gap $\Delta(x)$ measures the difference between suboptimal rate allocations and the optimal allocation.

III. Utility Gap and Stability

We first investigate how flow-level stability will be affected by utility gap. Consider a network where class-i flows arrive as a Poisson process of intensity $\lambda_i \geq 0$ and have i.i.d. exponential file sizes of mean $1/\mu_i$. Let $\rho_i = \lambda_i/\mu_i$ be the traffic intensity of class-i flows. This is the traffic load generated by class-i flows per unit time. Assume that flows remain in the network until they have transferred their given file size. The evolution of network state can be described by a Markov process $x(t)$, where $x_i(t)$ the number of class-i flows remaining in the system at time t. For a certain rate allocation policy $\phi(x)$, the transaction rates are given by

$$x_i \to x_i + 1, \quad \text{with rate } \lambda_i \quad (4)$$

$$x_i \to x_i - 1, \quad \text{with rate } \mu_i \phi_i(x) \quad (5)$$

We say the network is stable if the Markov process $x(t)$ is positive-concurrent, in which case the queue length does not blow up to infinity. If the feasible rate region $\mathcal{R}$ is compact and convex, a necessary stability condition has been given in [2], [3], [5]: The traffic intensity vector must belong to the interior of the feasible rate region ($\rho \in \check{\mathcal{R}}$). Furthermore, it has also been shown that all α-fair rate allocations with arbitrary $\alpha > 0$ maximize the flow-level stability region, i.e., assigning rates $\phi_{\mathrm{opt}}(x)$ achieves the maximum stability region, which is equal to $\check{\mathcal{R}}$. In other words, $\rho \in \check{\mathcal{R}}$ is a sufficient condition for stability with $\phi_{\mathrm{opt}}(x)$ as the optimal rate allocation policy.

In practice, when only suboptimal solutions are computable, a positive utility gap $\Delta(x) > 0$ exists. In the next section, we will derive a sufficient condition for achieving maximum stability. When the condition is not satisfied, we prove that the network is unstable and the achievable stability region may be strictly smaller than the feasible rate region. The main results on stability are stated in Theorem 1 and 2.

A. A Sufficient Condition for Maximum Stability

Theorem 1: For an arbitrary suboptimal rate allocation policy $\phi(x)$ and any positive $\alpha > 0$, if the order of the utility gap

$\Delta(x)$ in (3) is less than the order of the optimal utility when the number of active flows grows large, i.e.,

$$\lim_{\max_i x_i \to \infty} \frac{\Delta(x)}{\left|\sum_i x_i U_i(\frac{\phi_{\text{opt},i}}{x_i})\right|} = 0 \tag{6}$$

then the network is stable if the traffic condition $\rho \in \check{\mathcal{R}}$ is satisfied, i.e., the maximum stability region can be obtained.

Remark 1: If the utility gap $\Delta(x)$ is upper bounded by a constant for all states x, the maximum stability region can be achieved. This is simply a special case of Theorem 1 and can be easily proved by verifying utility gap condition (6). The statement holds for all α-fair utilities with $\alpha > 0$.

Remark 2: Theorem 1 shows that for achieving the maximum stability region, solving the optimal solution to the utility optimization problem (1) is not required. Thus, in practice, suboptimal rate allocation policies that may only require a much lower computational complexity or operate on a larger time-scale than that of the optimal policies could still stabilize the network, as long as the utility gap condition (6) is satisfied. The sufficient condition in Theorem 1 characterizes a class of suboptimal rate allocation policies that retain the maximum network stability.

B. A Lower Bound on Achievable Stability Region

When the condition (6) in Theorem 1 is not satisfied and the utility gap $\Delta(x)$ is on the same order as that of the optimal utility, the achievable stability region could be smaller than the feasible rate region.

Proposition 1: There exists a suboptimal rate allocation policy such that the utility gap is the same order as the order of the optimal utility, i.e., for some constant $\eta \in (0, 1)$,

$$\limsup_{\max_i x_i \to \infty} \frac{\Delta(x)}{\left|\sum_{i=1}^N x_i U_i(\frac{\phi_{\text{opt},i}}{x_i})\right|} \leq \eta, \tag{7}$$

but the achievable stability region is strictly smaller than $\check{\mathcal{R}}$, even if $\phi(x)$ is Pareto-optimal (i.e., $\phi(x)$ lies on the boundary of the feasible rate region).

Proposition 1 implies that there exists a suboptimal rate allocation policy whose achievable stability region is strictly smaller than the feasible rate region. Raised from this example, a challenge is to answer the question: what is the minimum stability region that a suboptimal rate allocation policy can achieve given that condition (7) is satisfied?

Theorem 2: For an arbitrary suboptimal rate allocation policy $\phi(x)$ and any positive $\alpha \neq 1$, if the order of the utility gap $\Delta(x)$ is the same as that of the optimal utility, i.e.,

$$\limsup_{\max_i x_i \to \infty} \frac{\Delta(x)}{\left|\sum_{i=1}^N x_i U_i(\frac{\phi_{\text{opt},i}}{x_i})\right|} \leq \eta \tag{8}$$

then the achievable stability region is lower bounded by $(1 - \eta)^{\frac{1}{\lceil 1-\alpha \rceil}} \check{\mathcal{R}}$. There also exists a suboptimal rate allocation policy whose stability region is exactly $(1-\eta)^{\frac{1}{\lceil 1-\alpha \rceil}} \check{\mathcal{R}}$, i.e., the lower bound is tight.

Remark 3: Theorem 2 provides a lower bound for achievable stability regions. Of course, under condition (8), there might still exist certain suboptimal rate allocation policies that are capable of achieving the maximum stability. However, the lower bound in Theorem 2 is tight in the sense that there exists a suboptimal rate allocation policy (see Equation (45)) and its stability region is exactly $(1 - \eta)^{\frac{1}{\lceil 1-\alpha \rceil}} \mathcal{R}$. Proposition 1 and Theorem 2 together characterize the stability of a broad class of suboptimal rate allocation policies.

C. Numerical Examples

Consider a simple network with routing matrix:

$$R = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \tag{9}$$

The linear network consists of $L = 2$ unit-capacity links with $x_0 > 0$ active flows on route 0, which crosses both links, and x_i active flows on route i, which uses link i alone, for $i = 1, 2$. For $\alpha > 0$, it is easy to compute the optimal rate allocation policy $\phi_{\text{opt}}(x)$ as follows:

$$\phi_{\text{opt},0}(x) = \frac{x_0}{x_0 + (x_1^\alpha + x_2^\alpha)^{\frac{1}{\alpha}}} \tag{10}$$

$$\phi_{\text{opt},i}(x) = 1 - \phi_{\text{opt},0}(x), \text{ for } i = 1, 2 \tag{11}$$

In order to obtain a suboptimal rate allocation policy $\phi(x)$ that satisfies the utility gap condition (6) in Theorem 1, we randomly perturb the optimal rate allocation at each state x, such that the utility gap remains bounded by a constant β, i.e.,

$$\Delta(x) = \sum_{i=1}^N x_i U_i(\frac{\phi_{\text{opt},i}}{x_i}) - \sum_{i=1}^N x_i U_i(\frac{\phi_i}{x_i}) \leq \beta \tag{12}$$

According to Remark 1, suboptimal rate allocation policy $\phi(x)$ achieves the maximum stability region that equals to the feasible rate region $\mathcal{R} = \{\phi \in \mathbb{R}_+^N : \phi_0 + \phi_1 \leq 1, \phi_0 + \phi_2 \leq 1\}$.

Figure 1 examines the stability of the suboptimal rate allocation policy $\phi(x)$ by plotting the average total queue length under the suboptimal rate allocation policy, with varying traffic load. The arrivals to the three flow classes are independent, Poisson distributed with equal traffic intensity $\rho_i = \rho_0$ for $i = 1, 2$. We assume proportional fairness of $\alpha = 1$. The curve under the suboptimal policy $\phi(x)$ approaches that under the optimal policy when the utility gap decreases.

IV. Utility Gap and Network Performance

Section III showed that when the utility gap is small enough, the stability of networks will remain unaffected. Therefore, a suboptimal rate allocation policy that under-optimizes the utility may still achieve the maximum stability region. On the other hand, since α-fair utility functions are designated for fairness objectives, it is clear that allowing a utility gap (or, equivalently, under-optimizing the utility) gives us freedom to potentially improve other network performance metrics, such as total throughput and maximum link saturation. Thus there exists a tradeoff between utility gap and the maximum network performance improvement we can potentially achieve. In this section we first provide a formulation of this tradeoff, then

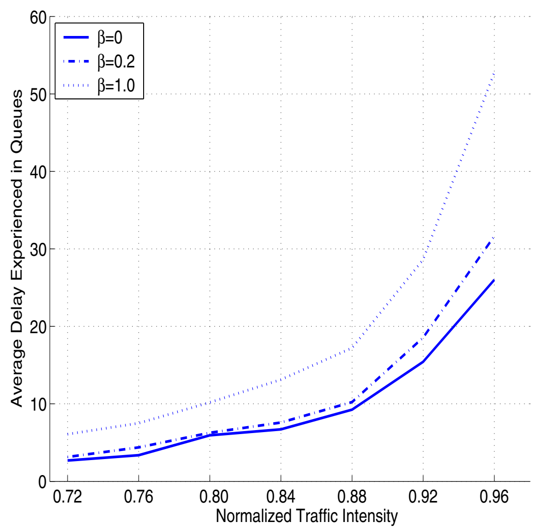

Fig. 1. This figure plots the average delay characteristics of three rate allocation policies corresponding to different β. It is shown that the two suboptimal rate allocation policies with $\beta > 0$ still stabilize the queue for normalized traffic intensity $\rho/c < 1$, although their delay performances in terms of the average queue length are worse compared to that of the optimal rate allocation policy with $\beta = 0$.

develop a quantitative approximation of the tradeoff curve based on local sensitivity analysis, assuming the utility gap is small. Our results in this section answer the following question related to utility gap: what is the maximum performance improvement we can possibly achieve by under-optimizing an α-fair utility with gap Δ?

We consider a major special case of the network model described in section III and define the feasible rate regions of wireline networks as follows. Consider a network of L links, indexed by l, each with a finite link capacity c_l. It is shared by N flow classes. Again we use ϕ_i to denote the total data rate of class-i flows. Then the feasible rate regions are defined by $\mathcal{R} = \{\phi : R\phi \preceq c\}$, where c is the vector of link capacities and R is the $L \times N$ routing matrix: $R_{li} = 1$ if class-i flows uses link l and 0 otherwise. At each state x, the optimal rate allocation is obtained by solving problem (1) with α-fair utility, i.e.,

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{N} x_i^{\alpha} \frac{\phi_i^{1-\alpha}}{1-\alpha} && (13) \\ \text{subject to} \quad & R\phi \preceq c, \;\; \phi \succeq 0 && (14) \\ \text{variables} \quad & \phi \end{aligned}$$

Let ϕ_{opt} be the optimal rate allocation that solves the maximization problem (13). Any suboptimal rate allocation $\phi \neq \phi_{opt}$ only achieves a utility value less than the maximum utility. We say a rate allocation ϕ under-optimizes the α-fair utility by a gap Δ if

$$\Delta = U_{\text{opt}} - \sum_{i=1}^{N} x_i^{\alpha} \frac{\phi_i^{1-\alpha}}{1-\alpha} \tag{15}$$

where $U_{\text{opt}} = \sum_{i=1}^{N} x_i^{\alpha} \phi_{\text{opt},i}^{1-\alpha}/(1-\alpha)$ is the optimal utility achieved by rate allocation ϕ_{opt}. Since the α-fair utility is designated for achieving fairness, under-optimizing the α-fair utility with a gap Δ relaxes the maximization problem (13). Thus it gives freedom to system designers to potentially improve other network performance objectives, such as total throughput and maximum link saturation. However, it is unclear how much performance improvement we can achieve by under-optimizing the α-fair utility with a given allowable gap. For example, if we prepare to sacrifice 5% of the utility, how much is the throughput improvement we could expect in return? In the next section, we formulate this type of tradeoff functions and provide a local sensitivity analysis based on examining the Karush-Kuhn-Tucker (KKT) conditions at the optimum allocation.

A. Utility Gap and Total Throughput

First we consider the tradeoff between utility gap and total throughput. For a rate allocation policy ϕ, total throughput T is simply defined as the sum-rate of all flow classes:

$$T = \sum_{i=1}^{N} \phi_i. \tag{16}$$

We are interested in characterizing the tradeoff between the utility gap and the maximum total throughput. More precisely, we compute the maximum total throughput that can be achieved by under-optimizing the utility with a designated gap. Thus maximum total throughput T is formulated as a function of the utility gap Δ. With some abuse of notation, we refer to this tradeoff function as $T(\Delta)$, which is defined as the maximized objective value of the following optimization problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{N} \phi_i && (17) \\ \text{subject to} \quad & R\phi \preceq c, \;\; \phi \succeq 0 \\ & \sum_{i=1}^{N} x_i^{\alpha} \frac{\phi_i^{1-\alpha}}{1-\alpha} \geq U_{\text{opt}} - \Delta \\ \text{variables} \quad & \phi. \end{aligned}$$

Thus gap Δ is the input and function $T(\Delta)$ gives the maximum possible total throughput under the utility gap constraint.

Remark 4: For the tradeoff function defined by (17), it is easy to see that increasing utility gap relaxes the constraint set of the optimization problem, and leads to a higher optimal objective value. Maximum total throughput $T(\Delta)$, defined by the optimization in (17), is a monotonically increasing function of the utility gap Δ.

Since the α-fair utility functions are concave and the total throughput is linear, we conclude that the maximization problem (17) is convex. Thus we can numerically solve it and compute the maximum-throughput-versus-utility tradeoff curve using any convex optimization solvers. Furthermore, according to the results in section III, a suboptimal rate allocation policy with small enough utility gap can still retain the maximum stability region. When the utility gap is small,

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2008 proceedings.

we can quantitatively approximate the maximum-throughput-versus-utility tradeoff function using its first order expansion:

$$T - T_0 = \left[\frac{dT}{d\Delta}\Big|_{\Delta=0} \right] \Delta + o(\Delta) \qquad (18)$$

where $T_0 = \sum_{i=1}^{N} \phi_{\mathrm{opt},i}$ is the total throughput at $\Delta = 0$.

In the context of convex optimization, the first order derivative $dT/d\Delta$, also known as shadow price, can be obtained by a local sensitivity analysis, if we make the assumption that the active constraint set in the problem (17) is unchanged or is perturbed locally, so that routing matrix R is fixed and independent of the input variable Δ. Further, we assume that the routing matrix R consists only of 'bottleneck' links. These two conditions guarantee that the tradeoff function $T(\Delta)$ is continuous and differentiable at a given Δ. These local sensitive analysis can provide a good approximation of the maximum-throughput-versus-utility tradeoff, when the α-fair utility is slightly under-optimized.

Theorem 3: The maximum-throughput-versus-utility tradeoff function $T(\Delta)$ has the following first order gradient (shadow price) at $\Delta = 0$:

$$\frac{dT}{d\Delta}\Big|_{\Delta=0} = -\frac{\mathbf{1}^T \cdot A \cdot \left(\frac{x^\alpha}{\phi_{\mathrm{opt}}^\alpha}\right)}{\left(\frac{x^\alpha}{\phi_{\mathrm{opt}}^\alpha}\right)^T \cdot A \cdot \left(\frac{x^\alpha}{\phi_{\mathrm{opt}}^\alpha}\right)}, \qquad (19)$$

where $A = D^{-1} - D^{-1}R^T \left(RD^{-1}R^T\right)^{-1} RD^{-1}$ and $D = \alpha \cdot \mathrm{diag}\left\{[\phi_{\mathrm{opt},1}^{-\alpha-1}; \ldots; \phi_{\mathrm{opt},N}^{-\alpha-1}]\right\}$ is a diagonal matrix. The vector division and power $x^\alpha/\phi_{\mathrm{opt}}^\alpha$ are component-wise.

B. Utility Gap and Maximum Link Saturation

In this section, we consider the maximal link saturation as a network performance metric, defined by

$$Z = \max_{l \in L} \frac{\sum_i R_{il}\phi_i}{c_l}. \qquad (20)$$

By under-optimizing the α-fair utility, it is possible to reduce the maximal link saturation and then balance the network traffic over all links. Moreover, reducing Z could potentially minimize the occurrence of 'bottleneck' links in the network, and also make the network more robust to link capacity fluctuation and traffic bursts.

We characterize the optimal tradeoff between the utility gap and the maximum link saturation, i.e., we compute the minimum Z that can be achieve by under-optimizing the α-fair utility with a designated gap. This tradeoff function $Z(\Delta)$ can be formulated as follows

$$\begin{aligned} Z(\Delta) = \quad & \text{minimize} \quad && \max_{l \in L} \frac{\sum_s R_{il}\phi_i}{c_l} \\ & \text{subject to} \quad && R\phi \preceq c, \ \phi \succeq 0 \\ & && \sum_{i=1}^{N} x_i^\alpha \frac{\phi_i^{1-\alpha}}{1-\alpha} \geq U_{\mathrm{opt}} - \Delta \\ & \text{variables} \quad && \phi \end{aligned} \qquad (21)$$

Remark 5: For the tradeoff function defined by (21), it is easy to see that increasing utility gap relaxes the constraint set of the optimization problem, and leads to a smaller optimal objective value. Thus maximum link saturation Z is a monotonically decreasing function of the utility gap Δ. Furthermore, it is easy to verify that the optimization problem (21) is convex. The saturation-gap tradeoff can be numerically computed.

Now we conduct a local sensitivity analysis for the saturation-gap tradeoff defined by optimization problem (21). Again, we make the assumption that the active constraint set in the problem (17) is unchanged or is perturbed locally, so that routing matrix R is fixed and independent of the input variable Δ. The main result is summarized in the next theorem. Its proof is very similar to that of Theorem 3. Let Z_0 be the link saturation at $\Delta = 0$. We have $Z_0 = 1$ since the rate allocation policy ϕ_{opt} must be Pareto-optimal at $\Delta = 0$.

Theorem 4: When the utility gap is small, the saturation-utility tradeoff function can be approximated using its first order expansion:

$$Z - Z_0 = \left[\frac{dZ}{d\Delta}\Big|_{\Delta=0} \right] \Delta + o(\Delta). \qquad (22)$$

The first order derivative (shadow price) of the saturation-utility tradeoff function is given by

$$\frac{dZ}{d\Delta}\Big|_{\Delta=0} = -\frac{1}{c^T \left(RD^{-1}R^T\right)^{-1} c}, \qquad (23)$$

where $D = \alpha \cdot \mathrm{diag}\left\{[x_1^\alpha \phi_{\mathrm{opt},1}^{-\alpha-1}, \ldots, x_N^\alpha \phi_{\mathrm{opt},N}^{-\alpha-1}]\right\}$ is diagonal.

C. Numerical Examples

In this section, we plot the two tradeoff curves and their gradient approximations obtained in Section IV.A and Section IV.B, for the linear network in Section III.C. Assume all links have unit capacity of $c_i = 1$.

Let x_i denote the number of active flows for source i. We can solve the two convex optimization problems (17) and (21) for an arbitrary gap Δ and thus obtain the exact tradeoff curves $T(\Delta)$ and $Z(\Delta)$, which are plotted in Figure 2 and Figure 3 in solid line. In both figures, we assume that the number of active flows are $x_0 = 10$ and $x_i = 5$ for $i = 1, 2$. The proportional fairness utility with $\alpha = 1$ is considered.

When the utility gap Δ is small, the maximum-throughput-versus-utility and the saturation-gap tradeoff curves can be approximated by their first order expansions given by (18) and (22) respectively. Using the close form solutions in Theorem 3 and Theorem 4, we compute the first order gradients as follows $\frac{dT}{d\Delta}\Big|_{\Delta=0} = 1.1899$ and $\frac{dZ}{d\Delta}\Big|_{\Delta=0} = -0.0697$. Thus the two tradeoff curves can be approximated by

$$T(\Delta) \approx T_0 + 1.1899\Delta \qquad (24)$$

$$Z(\Delta) \approx Z_0 - 0.0697\Delta \qquad (25)$$

In Figure 2 and Figure 3, we also plot the corresponding linear approximations for the maximum-throughput-versus-gap and the saturation-gap tradeoff curves in dashed line.

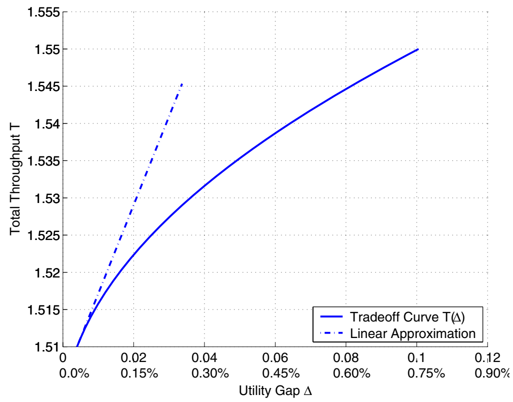

Fig. 2. Throughput-Gap tradeoff curve and its first order approximation.

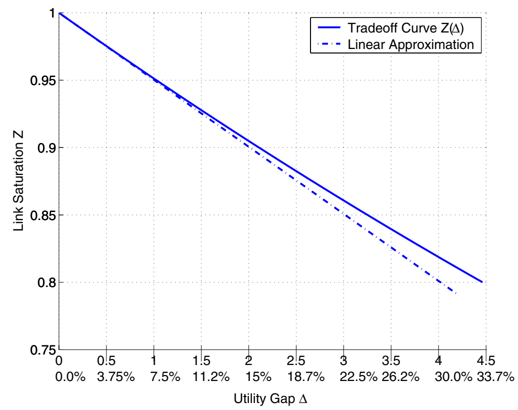

Fig. 3. Saturation-Gap tradeoff curve and its first order approximation.

Figure 3 shows that the saturation-gap tradeoff defined in Section IV.B can be well approximated by its first order expansion, given by the closed-form expressions in Theorem 4, while such an approximation is accurate for the throughput-gap tradeoff only when the utility gap is very small (Figure 2). These two tradeoff curves indicates the performance improvement we can achieve by under-optimizing the utility with a designated small utility gap. For example, if we under-optimize the utility by 3‰, i.e., $\Delta = 3‰|U_{\text{opt}}| = 0.04$, it is clear from equation (24) and equation (25) that a maximum total throughput increase of $T - T_0 = +0.05$ (equivalently $+3.3\%T_0$) and a link saturation reduction of $Z - Z_0 = -0.003$ (equivalently $-0.3\%Z_0$) could be expected in return. This result not only illustrates the potential benefits of under-optimizing an α-fair utility, but also quantitatively characterizes the tradeoff between sacrificing utility value and achieving network performance improvement. Whether this particular tradeoff is worth making or not depends on operator's preference, but it is important to provide the choices of tradeoff through the results like those in this section.

V. Concluding Remarks

Suboptimal resource allocation with a utility gap is simply an inevitable phenomenon in practical networking. Fortunately, it may still be able to maintain stability region and even enhance other network performance metrics. Intuition on stability and utility–versus-throughput and utility-versus-saturation tradeoff are quantified with closed-form expressions in this paper. There are still open questions to the study of suboptimal solutions to network optimization, e.g., degradation on fairness due to utility gap and global sensitivity analysis, before we fully understand "how bad is suboptimal rate allocation".

Appendix: Proofs

A. Proof of Theorem 1.

Proof: The network is stable if the Markov process $x(t)$ is positive-recurrent. According to the Foster's criterion [4], this is true if for some finite set $A \subset \mathbb{Z}_+^N$, some non-negative function f and some positive constant $\kappa > 0$,

$$\forall x \notin A,\ df(x) = \sum_{y \neq x} q(x,y)\,[f(y) - f(x)] \leq -\kappa \quad (26)$$

where $q(x,y)$ is the transition rate from state x to state y. Toward this end, we first prove an important lemma for all suboptimal rate allocations satisfying (6).

Lemma 1: Consider any traffic intensity $\rho \in \check{\mathcal{R}}$. There exist positive constants ϵ and γ, such that for any suboptimal rate allocation $\phi(x)$ satisfying (6) and any network state satisfying $\max_i x_i/\rho_i > \gamma$, the following inequality holds:

$$\sum_{i:x_i \geq 1} x_i^{\alpha} \rho_i^{-\alpha} \left[(1+\epsilon)\rho_i - \phi_i\right] \leq 0 \quad (27)$$

Proof: First, we consider the case that $0 < \alpha < 1$ and thus the optimal utility is positive. Under the traffic condition $\rho \in \check{\mathcal{R}}$, there exist $\epsilon \geq 0$ and $\delta \geq 0$ such that rate vector $(1+\epsilon)(1-\delta)^{-\frac{1}{1-\alpha}}\rho$ satisfies the feasible rate constraints, i.e.,

$$(1+\epsilon)(1-\delta)^{-\frac{1}{1-\alpha}}\rho \in \mathcal{R} \quad (28)$$

According to the condition (6), we can conclude that there exists a positive γ such that for $\max_i x_i/\rho_i > \gamma$ and $\delta \geq 0$,

$$\Delta(x) \leq \delta \left| \sum_i x_i U_i\left(\frac{\phi_{\text{opt},i}}{x_i}\right) \right| \quad (29)$$

Let $u \in \mathcal{R}$ be an arbitrary rate vector and $\delta_0 = (1-\delta)^{\frac{1}{1-\alpha}}$. Since the optimal utility is positive, in view of (3) we obtain the following inequalities:

$$\begin{aligned} 0 &\le \sum_{i=1}^{N} x_i U_i(\frac{\phi_i}{x_i}) + \Delta(x) - \sum_i x_i U_i(\frac{\phi_{\text{opt},i}}{x_i}) \\ &\le \sum_{i=1}^{N} x_i U_i(\frac{\phi_i}{x_i}) - (1-\delta)\sum_i x_i U_i(\frac{\phi_{\text{opt},i}}{x_i}) \\ &\le \sum_{i=1}^{N} x_i U_i(\frac{\phi_i}{x_i}) - (1-\delta)\sum_i x_i U_i(\frac{u_i}{x_i}) \\ &= \sum_{i=1}^{N} \left[(\phi_i - \delta_0 u_i) U_i'(\frac{\delta_0 u_i}{x_i}) + \frac{(\phi_i - \delta_0 u_i)^2}{2x_i} U_i''(\frac{y_i}{x_i})\right] \\ &\le \sum_{i=1}^{N} \left[\phi_i - \delta_0 u_i\right] U_i'(\frac{\delta_0 u_i}{x_i}) \end{aligned}$$

where y_i is a proper scalar between $\delta_0 u_i$ and ϕ_i. The third step follows since ϕ_{opt} is the optimal rate allocation that maximized the α-fair utility. The forth equation is from the Taylor's Formula and the last step is due to the fact that utility function $U_i(\cdot)$ is concave. In view of (2), this becomes

$$\sum_{i=1}^{N} x_i^{\alpha} \frac{(\delta_0 u_i - \phi_i)}{\delta_0 u_i^{\alpha}} \le 0 \tag{30}$$

Choosing $u = \frac{1+\epsilon}{\delta_0}\rho$ and applying the previous inequality, we get

$$\sum_{i=1}^{N} x_i^{\alpha} \rho_i^{-\alpha} \left[(1+\epsilon)\rho_i - \phi_i\right] \le 0 \tag{31}$$

This is exactly (27) since flow classes with zero flow don't contribute to the summation. Similarly, when $\alpha > 1$ and the optimal utility is negative, there exist $\epsilon \ge 0$ and $\delta \ge 0$ such that rate vector $(1+\epsilon)(1+\delta)^{\frac{1}{\alpha-1}}\rho$ satisfies the feasible rate constraints. Using the same proof technique and choosing $u = (1+\epsilon)(1+\delta)^{\frac{1}{\alpha-1}}\rho$, we can show that condition (31) is also satisfied when $\alpha > 1$ and $\max_i x_i/\rho_i > \gamma$. For $\alpha = 1$, choosing $u = [(1+\epsilon)\rho]^{\frac{1}{1-\delta}}$ leads to the same result. Thus the lemma holds for any positive $\alpha > 0$. ■

Now we prove the Foster's drift condition. For a suboptimal rate allocation policy $\phi(x)$, transition rates of the network states are given by equation (4) and (5). We consider a positive function $f(x)$ defined on the set of all network states by

$$f(x) = \sum_{i=1}^{N}\sum_{n=1}^{x_i} \mu_i^{-1} \frac{n^{\alpha}}{\rho_i^{\alpha}} \tag{32}$$

Then we obtain

$$\begin{aligned} df(x) &= \sum_{i=1}^{N}\left[\frac{\lambda_i}{\mu_i}\frac{(x_i+1)^{\alpha}}{\rho_i^{\alpha}} - \phi\frac{x_i^{\alpha}}{\rho_i^{\alpha}}\right] \qquad (33) \\ &= \sum_{i:x_i\ge 1} \frac{x_i^{\alpha}}{\rho_i^{\alpha}}\left[\left(1+\frac{1}{x_i}\right)^{\alpha}\rho_i - \phi_i\right] + \sum_{i:x_i=0}\rho_i^{1-\alpha} \end{aligned}$$

Given $\alpha > 0$, it is easy to see that there exists a positive constant $K_\alpha > 0$ such that for all $x_i \ge 1$,

$$\left(1+\tfrac{1}{x_i}\right)^{\alpha} \le 1 + K_\alpha \frac{1}{x_i} \tag{34}$$

Plugging this result into (33), we obtain:

$$df(x) \le \sum_{i:x_i\ge 1}\frac{x_i^{\alpha}}{\rho_i^{\alpha}}\left[\left(1+\frac{K_\alpha}{x_i}\right)\rho_i - \phi_i\right] + \sum_{i:x_i=0}\rho_i^{1-\alpha}$$

Let n =argmax$_i$ x_i/ρ_i. In view of Lemma 1, for all network states satisfying $x_n/\rho_n \ge \gamma$, the drift $df(x)$ is bounded by

$$\begin{aligned} df(x) &\le \sum_{i:x_i\ge 1}\frac{x_i^{\alpha}}{\rho_i^{\alpha}}\left(K_a\frac{\rho_i}{x_i} - \rho_i\epsilon\right) + \sum_{i:x_i=0}\rho_i^{1-\alpha} \\ &\le -\rho_n\epsilon\frac{x_n^{\alpha}}{\rho_n^{\alpha}} + \sum_{i:x_i\ge 1}\frac{x_i^{\alpha}}{\rho_i^{\alpha}}\left(K_a\frac{\rho_i}{x_i}\right) + \sum_{i:x_i=0}\rho_i^{1-\alpha} \\ &\le -\rho_0\epsilon\frac{x_n^{\alpha}}{\rho_n^{\alpha}} + NK_\alpha\frac{x_n^{\alpha-1}}{\rho_n^{\alpha-1}} + \sum_{i:x_i=0}\rho_i^{1-\alpha} \\ &\le \frac{x_n^{\alpha-1}}{\rho_n^{\alpha-1}}\left(NK_\alpha - \epsilon\rho_0\frac{x_n}{\rho_n}\right) + \sum_{i:x_i=0}\rho_i^{1-\alpha} \end{aligned}$$

where ρ_0 denotes the minimum component of traffic intensity ρ. With straight forward computation, it is easy to show that $df(x) \to -\infty$ as $x_n/\rho_n \to +\infty$. Thus there exists positive constants ξ and κ, such that for $x_n/\rho_n > \xi$, we have $df(x) < -\kappa$.

We define the set A in the Foster's criteria as follows

$$A \triangleq \left\{x \in \mathbb{R}_+^N : x \le \rho\cdot\max(\gamma,\xi)\right\} \tag{35}$$

Then we deduce that the negative draft condition holds for all $x \notin A$, i.e., $df(x) < -\kappa,\quad$ for all $x \notin A$. The process $x(t)$ is positive-concurrent. This implies that the suboptimal rate allocation policy $\phi(x)$ stabilizes the network when the utility gap condition (6) is satisfied. ■

B. Proof of Proposition 1.

Proof: We construct a counter-example. Consider a network with two classes of users and a feasible rate region depicted in Figure 4. For $\alpha = 1/2$, let $\phi_{\text{opt}}(x)$ denote the

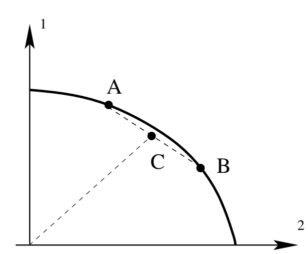

Fig. 4. The feasible rate region under consideration.

optimal rate allocation at state x. We define a suboptimal rate allocation by

$$\phi(x) = \begin{cases} \phi_{\text{opt}}(x), & \text{if } \phi_{\text{opt}} \text{ doesn't lie on } \widehat{AB} \\ \phi_A, & \text{otherwise, if } x_1 > x_2 \\ \phi_B, & \text{otherwise, if } x_1 \le x_2 \end{cases} \tag{36}$$

where $\widehat{AB}$ denotes the boundary of the rate region between point A and B, and ϕ_A and ϕ_B are the rates at point A and B respectively. First, we show that the gap is on the same order as that of the optimal utility, i.e.,

$$\begin{aligned}
&\limsup_{\max_i x_i \to \infty} \frac{\Delta(x)}{\left|\sum_{i=1}^N x_i U_i(\frac{\phi_{\text{opt},i}}{x_i})\right|} \\
&= \left(1 - \limsup_{\max_i x_i \to \infty} \frac{\sqrt{x_1\phi_1} + \sqrt{x_2\phi_2}}{\sqrt{x_1\phi_{\text{opt},1}} + \sqrt{x_2\phi_{\text{opt},2}}}\right) \cdot \mathbf{1}_{\{\phi_{\text{opt}} \in AB\}} \\
&\leq 1 - \limsup_{\max_i x_i \to \infty} \frac{\sqrt{x_1\phi_{B,1}} + \sqrt{x_2\phi_{A,2}}}{\sqrt{x_1\phi_{A,1}} + \sqrt{x_2\phi_{B,2}}} \\
&\leq 1 - \max\left(\sqrt{\frac{\phi_{B,1}}{\phi_{A,1}}}, \sqrt{\frac{\phi_{A,2}}{\phi_{B,2}}}\right)
\end{aligned}$$

Therefore, condition (8) is satisfied.

Next, we show that for small enough $\epsilon > 0$, the network is unstable under traffic intensity $\rho = (1+\epsilon)\phi_C$. To prove this, we define a Lyapunov function by

$$f(x) = \mu_1^{-1} w_1 x_1 + \mu_2^{-1} w_2 x_2 \tag{37}$$

where $w_1 = \phi_{B,2} - \phi_{A,2}$ and $w_2 = \phi_{A,1} - \phi_{B,1}$ are positive. Then we prove that the drift is always positive for traffic intensity $\rho = (1+\epsilon)\phi_C$ with a small enough $\epsilon > 0$, i.e.,

$$\begin{aligned}
df(x) &= \sum_{x \neq y} q(x,y)\left[f(y) - f(x)\right] \\
&= \epsilon(w_1\phi_{C,1} + w_2\phi_{C,2}) + w_1(\phi_{C,1} - \phi_1) \\
&\quad + w_2(\phi_{C,2} - \phi_2) \\
&\geq \epsilon(w_1\phi_{C,1} + w_2\phi_{C,2}) > 0
\end{aligned}$$

where the third step holds since the suboptimal rate allocation $\phi(x)$ always lies below the line AB. Thus, for the choice of Lyapunov function (37), the drift is always positive. This implies that the network is unstable. ■

C. Proof of Theorem 2.

Proof: We use the same proof technique as in Theorem 1, and show that provided the traffic condition is satisfied, the network is stable.

Lemma 2: Consider any traffic intensity $\rho \in (1-\eta)^{\frac{1}{1-\alpha}}\mathcal{R}$. There exist positive constants ϵ and γ, such that for any suboptimal rate allocation $\phi(x)$ satisfying (8) and any network state satisfying $\max_i x_i/\rho_i > \gamma$, the following inequality holds:

$$\sum_{i: x_i \geq 1} x_i^\alpha \rho_i^{-\alpha}\left[(1+\epsilon)\rho_i - \phi_i\right] \leq 0 \tag{38}$$

Proof: First, we consider the case that $\alpha < 1$ and thus the optimal utility is positive. Under the traffic condition $\rho \in (1-\eta)^{\frac{1}{1-\alpha}}\check{\mathcal{R}}$, there exist $\epsilon \geq 0$ and $\delta \geq 0$ such that rate vector $(1+\epsilon)[1-(1+\delta)\eta]^{-\frac{1}{1-\alpha}}\rho$ satisfies the feasible rate constraints, i.e.,

$$(1+\epsilon)[1-(1+\delta)\eta]^{-\frac{1}{1-\alpha}}\rho \in \mathcal{R} \tag{39}$$

According to the condition (8), we can conclude there exists a positive γ such that for $\max_i x_i > \gamma$ and $\delta \geq 0$,

$$\Delta(x) \leq \eta(1+\delta)\left|\sum_{i=1}^N x_i U_i(\frac{\phi_{\text{opt},i}}{x_i})\right| \tag{40}$$

Let $u \in \mathcal{R}$ be an arbitrary rate vector and $\delta_0 = [1-(1+\delta)\eta]^{\frac{1}{1-\alpha}}$. Since the optimal utility is positive, we obtain

$$\begin{aligned}
0 &\leq \sum_{i=1}^N X_i U_i(\frac{\phi_i}{X_i}) + \Delta(X) - \sum_{i=1}^N X_i U_i(\frac{\phi_{\text{opt},i}}{X_i}) \\
&\leq \sum_{i=1}^N X_i U_i(\frac{\phi_i}{X_i}) - [1-(1+\delta)\eta]\sum_{i=1}^N X_i U_i(\frac{u_i}{X_i}) \\
&\leq \sum_{i=1}^N \left[\phi_i - \delta_0 u_i\right] U_i'(\frac{\delta_0 u_i}{X_i})
\end{aligned}$$

where y_i is a proper scalar between $\delta_0 u_i$ and ϕ_i and the last step follows since the α-fair utility function $U_i(\cdot)$ is concave. In view of (2), the above inequality can be rewritten as

$$\sum_{i=1}^N X_i^\alpha \frac{(\delta_0 u_i - \phi_i)}{u_i^\alpha} \leq 0 \tag{41}$$

Choosing $u = \frac{1+\epsilon}{\delta_0}\rho \in \mathcal{R}$, we obtain

$$\sum_{i=1}^N X_i^\alpha \rho_i^{-\alpha}\left[(1+\epsilon)\rho_i - \phi_i\right] \leq 0 \tag{42}$$

This is exactly (38) since flow classes with zero active flows are allocated rate zero.

Similarly, when $\alpha > 1$, there exist $\epsilon \geq 0$ and $\delta \geq 0$ such that rate vector $(1+\epsilon)[1+(1+\delta)\eta]^{\frac{1}{\alpha-1}}\rho$ satisfies the feasible rate constraints. Using the same proof technique and choosing $u = (1+\epsilon)[1+(1+\delta)\eta]^{\frac{1}{\alpha-1}}\rho$, we can show that condition (42) is also satisfied for $\alpha > 1$ and $\max_i x_i > \gamma$. ■

To prove stability, again we consider the Lyapunov function defined by

$$f(x) = \sum_{i=1}^N \sum_{n=1}^{x_i} \mu_i^{-1}\frac{n^\alpha}{\rho_i^\alpha} \tag{43}$$

Using the results in the proof of Theorem 1, we conclude that there exist positive constants ξ and κ, such that for $\max_i(x_i/\rho_i) > \max(\xi, \gamma)$, the negative drift condition holds:

$$\begin{aligned}
df(x) &= \sum_{i=1}^N \left[\lambda_i\mu_i\frac{(x_i+1)^\alpha}{\rho_i^\alpha} - \phi\frac{x_i^\alpha}{\rho_i^\alpha}\right] \\
&\leq \frac{x_n^{\alpha-1}}{\rho_n^{\alpha-1}}\left(NK_\alpha - \epsilon\rho_0\frac{x_n}{\rho_n}\right) + \sum_{i:x_i=0}\rho_i^{1-\alpha} \\
&\leq -\kappa
\end{aligned} \tag{44}$$

Then the Foster's criteria is satisfied on set $A \triangleq \{x \in \mathbb{R}_+^N : x \leq \rho \cdot \max(\gamma, \xi)\}$. The suboptimal rate allocation policy $\phi(x)$ stabilizes the network given that the utility gap condition (6) is satisfied.

To show that there exists a suboptimal rate allocation policy whose stability region is exactly $(1-\eta)^{\frac{1}{1-\alpha}}\mathcal{R}$, we assume $\alpha < 1$ and construct a suboptimal policy as follows

$$\phi_i(x) = (1-\eta)^{\frac{1}{1-\alpha}}\phi_{\text{opt},i}(x) \tag{45}$$

Then we compute the utility gap as

$$\begin{aligned}\Delta(x) &= \sum_i x_i U_i(\frac{\phi_{\text{opt},i}}{x_i}) - \sum_i x_i U_i(\frac{\phi_i}{x_i}) \\ &= \eta \sum_i x_i U_i(\frac{\phi_{\text{opt},i}}{x_i})\end{aligned} \tag{46}$$

Therefore, the utility gap is exactly η proportion of the optimal utility. Condition (8) is satisfied. Furthermore, it is not difficult to see that rate allocation $\phi(x)$ actually maximizes the utility on a reduced rate region $(1-\eta)^{\frac{1}{|1-\alpha|}}\mathcal{R}$. The achievable stability region is indeed $(1-\eta)^{\frac{1}{|1-\alpha|}}\mathcal{R}$. The proof for $\alpha > 1$ is similar and will be omitted here. We complete the proof of Theorem 2. ■

D. Proof of Theorem 3.

Proof: As the first step, we form the Lagrangian for the optimization problem (17) as

$$\mathcal{L}(\phi, p, q) = \sum_{i=1}^{N} \phi_i + p^T(c - R\phi) + q(V(\phi) + \Delta - U_{\text{opt}})$$

where $V(\phi) = \sum_{i=1}^{N} x_i^{\alpha}\phi_i^{1-\alpha}/(1-\alpha)$ is the achievable utility of a rate allocation ϕ. Vector p and scalar q are Lagrangian multipliers for the two constraints in (17) respectively. At the optimal point of (17), the KKT conditions for optimality are given by

$$R\phi = c, \quad V(\phi) = U_{\text{opt}} - \Delta \tag{47}$$

$$R^T p - q\frac{dV(\phi)}{d\phi} - \frac{d(\sum_{i=1}^{N}\phi_i)}{d\phi} = 0 \tag{48}$$

From the implicit function theorem, variables ϕ, p and q can be viewed as implicit functions of c and Δ, which are uniquely defined by the KKT conditions (47) and (48). We define two vectors $y = [\phi; q; p]$, $w = [c; \Delta]$ and a residual

$$G(y, w) = \begin{pmatrix} R\phi - c \\ U_{\text{opt}} + \Delta - V(\phi) \\ R^T p - q\frac{dV(\phi)}{d\phi} - \mathbf{1} \end{pmatrix} \tag{49}$$

where $\mathbf{1}$ is a $N \times 1$ vector consisting of all one's. Then the KKT conditions can be rewritten as $G(y, w) = 0$. The first order derivative of the residual $G(y, w)$ can be obtained as follows

$$\frac{\partial G}{\partial y} = \begin{pmatrix} R & \mathbf{0} & \mathbf{0} \\ -\frac{dV(\phi)}{d\phi} & \mathbf{0} & \mathbf{0} \\ qD & -\frac{dV(\phi)}{d\phi} & R^T \end{pmatrix} = \begin{pmatrix} \hat{R} & \mathbf{0} \\ qD & \hat{R}^T \end{pmatrix} \tag{50}$$

and

$$\frac{\partial G}{\partial w} = \begin{pmatrix} -I & \mathbf{0} \\ \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{51}$$

where $\hat{R} = [R; -dV(\phi)/d\phi]$ is an extended routing matrix and D is a diagonal matrix of the form

$$D = \alpha \cdot \text{diag}\left\{[\phi_1^{-\alpha-1}; \ldots; \phi_N^{-\alpha-1}]\right\} \tag{52}$$

Let $e = [0; \ldots; 0; 1]$. From the implicit function theorem, we obtain

$$\begin{aligned}\frac{dy}{dw} &= -\left(\frac{\partial G}{\partial y}\right)^{-1}\frac{\partial G}{\partial w}, \\ &= \begin{pmatrix} \hat{R} & \mathbf{0} \\ qD & \hat{R}^T \end{pmatrix}^{-1} \cdot \begin{pmatrix} -I & \mathbf{0} \\ \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \\ &= \begin{pmatrix} * & D^{-1}\hat{R}^T\left(\hat{R}D^{-1}\hat{R}^T\right)^{-1} e \\ * & * \\ * & * \end{pmatrix}\end{aligned} \tag{53}$$

This implies

$$\frac{\partial T}{\partial V} = \frac{d\left(\sum_{i=1}^{N}\phi_i\right)}{d\phi} \cdot \frac{d\phi}{d\Delta} = \mathbf{1}^T D^{-1}\hat{R}^T\left(\hat{R}D^{-1}\hat{R}^T\right)^{-1} e$$

Plugging the expression for $\hat{R}$ and performing some matrix manipulation, we derive the desired result. ■

References

[1] F.Kelly, A.Maulloo and D.Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237-252, 1998.

[2] T. Bonald and L. Massoulie, "Impact of Fairness on Internet Performance," *In Proceedings of ACM Sigmetrics*, pp. 82-91, Cambridge, MA, June 2001.

[3] H.Q. Ye, "Stability of Data Networks Under an Optimization-Based Bandwidth Allocation," *IEEE Transactions on Automatic Control*, vol. 48, no. 7, pp. 1238-1242, July 2003.

[4] T. Bonald, M. Massoulie, A. Proutiere and J. Virtamo, "A Queuing Analysis of Max-Min Fairness, Proportional Fairness and Balanced Fairness," *Special Issue of Queueing Systems: Queueing Models for Fair Resource Sharing*, vol. 53, pp. 65-84, June 2006.

[5] J. Liu, A. Proutiere, Y. Yi, M. Chiang and H. V. Poor, "Flow-Level Stability of Data Networks with Non-convex and Time-varying Rate Regions", *In Proceedings of ACM Sigmetrics*, pp. 239-250, June 2007.

[6] X. Lin and B. Shroff, "The Impact of Imperfect Scheduling on Cross-Layer Congestion Control in Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp 302-315, April 2006.

[7] X. Lin, N. B. Shroff and R. Srikant, "On the Connection Level Stability of Congestion-Controlled Communication Networks," *To appear in IEEE Transactions on Information Theory*.

[8] P. Giaccone, B. Prabhakar and D. Shah, "Towards simple, high-performance schedulers for high-aggregate bandwidth switches", *In Proceedings of IEEE Infocom*, pp. 1160-1169, New York City, June 2002.

[9] A. Eryilmaz, R. Srikant and J. Perkins, "Stable Scheduling Policies for Fading Wireless Channels", *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pages 411-424, April 2005.

[10] 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, October 2000.

[11] L. Massoulie and J. Roberts, "Bandwidth sharing: objectives and algorithms", *IEEE/ACM Transactions on Networking*, vol. 10, no. 3, pp. 320-328, June 2002.

[12] S. Kunniyur and R. Srikant, "End-to-end congestion control: utility functions, random losses and ECN marks", *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 689-702, October 2003.

[13] T. Bonald, S. Borst, N. Hegde and A. Proutiere, "Wireless data networks in multicell scenarios", *In Proceedings of ACM Sigmetrics*, pp. 378-380, June 2004

[14] A. Tang, J. Wang and S. Low, "Counter-intuitive Behaviors in Networks under End-to-end Control," *IEEE /ACM Transactions on Networking*, vol. 14, no. 2, pp. 355-368, April 2006.