Pricing a Shared Access Link for Fair and Efficient Operation with Variable User Data Rates

Joy Kuri and Sharmili Roy
Centre for Electronics Design and Technology
Indian Institute of Science
kuri@cedt.iisc.ernet.in sharmili@cedt.iisc.ernet.in

***Abstract*— The literature on pricing implicitly assumes an "infinite data" model, in which sources can sustain any data rate indefinitely. We assume a more realistic "finite data" model, in which sources occasionally run out of data; this leads to *variable* user data rates. Further, we assume that users have contracts with the service provider, specifying the rates at which they can inject traffic into the network. Our objective is to study how prices can be set such that a single link can be shared efficiently and fairly among users in a dynamically changing scenario where a subset of users occasionally has little data to send. User preferences are modelled by concave increasing utility functions. Further, we introduce two additional elements: a convex increasing *disutility* function and a convex increasing *multiplicative congestion-penalty function*. The disutility function takes the *shortfall* (contracted rate minus present rate) as its argument, and essentially encourages users to send traffic at their contracted rates, while the congestion-penalty function discourages heavy users from sending excess data when the link is congested. We obtain simple necessary and sufficient conditions on prices for fair and efficient link sharing; moreover, we show that a *single* price for all users achieves this. We illustrate the ideas using a simple experiment.**

I. Introduction and Related Work

Pricing has been suggested as a mechanism to control congestion and ensure fair and efficient operation of networks. In much of the published literature ([1] [2] [3] [4] [5], [6] [7]), elastic traffic is considered and the context of operation is as follows. Each user of the network has a utility function that quantifies the benefit that she derives from the network. The utility function is an increasing (often, concave increasing) function of the rate at which the user can send data through the network. The system objective is maximization of the sum of all users' utility functions. The problem is to find the vector of users' rates such that the system objective is realized.

The resulting constrained optimization problem can be solved in a centralized manner if all the utility functions were known. In practice, however, there is no central authority that computes rates and further, the users' utility functions are not known. In [1], Kelly proposed a decentralized method to arrive at the system-optimal rates. In this method, the network declares *prices*, and each user individually solves the problem of maximizing her *net benefit* or *net utility*, which is her utility minus the total cost paid to the network. He shows that there exists a vector of prices such that the vector of individually optimal rates arrived at by the users is, indeed, the system-optimal rate vector.

Nevertheless, several authors have pointed out that it is enough to work under the assumption of *direct revelation*, in which the users' utility functions are revealed to a central controlling authority ([8], [9]). In other words, we do not lose anything by assuming that the users' utility functions are known because there always exists an equivalent formulation based on direct revelation. Accordingly, we assume in this paper that users' utility functions are known to the central authority (the service provider) which can then use this knowledge to design appropriate prices. It will turn out that the complete utility function need not be known; actually, far less knowledge suffices — only the derivative of the utility function at a single point is enough. Further, [10] mentions that prices are used in two kinds of problems: one in which the objective is to promote fair and efficient resource sharing, and another in which the objective is maximization of the revenue earned by the central authority. As in [10], our objective in this paper is the former, *viz.,* fair and efficient sharing of a single access link; we do not consider the problem of revenue maximization.

The published literature, however, tacitly assumes an *infinite data model*. Every source is assumed to have an infinite backlog of data. The implication is that a source can send traffic at *any* rate (obtained from the solution to the individual optimization problem) continuously — there is never a dearth of data. In practice, of course, sources will occasionally run out of data. We consider a *finite* data model, in which, occasionally, a source does not have enough data to send. Therefore, a source may not be able to sustain a data rate that is suitable for a fair and efficient operation of the system.

Further, in practice, users have contracts with the service provider (SP) that specify the rates at which they can send traffic into the network. Hence, the practical problem is often *not* to find out the correct vector of rates by using a distributed algorithm—the correct vector is simply the vector of contracted rates. Rather, it is important to devise a scheme of operation such that any slack caused by a user who is sending traffic at a rate lower than her contracted rate—referred to as an "underuser" henceforth—can be utilized by others. Correspondingly, a user with plenty of data available is referred to as an "overuser" because she can send data at a rate higher than her contracted rate.

We also believe that one must have congestion-dependent and user-dependent pricing. If the network is not congested, then the price should remain low, so that users with excess data can utilize the network. But when the network becomes congested, the price should not increase equally for all users; rather, those users who have exceeded their contracted rates and have caused congestion should be charged heavily, while those who are compliant should be charged at no more than their nominal rates. Even though our framework allows different prices for different users, our analysis shows that a *single* price for all users suffices. This is attractive, because the management problem of maintaining prices for a possibly large number of users is solved very simply.

Summarizing, our approach is different from that in the literature in the following respects: *(a)* finite data model, *(b)* contracted rates and *(c)* congestion-dependent as well as user-dependent pricing.

We are interested in ensuring that network operation is characterised by the following.

- When some users are underusers because of limited available data, it should be possible for others to increase their rates so as to utilize the slack. Does there exist a pricing scheme such that users with plenty of data available are *encouraged* to become overusers? This means that in this situation, these users' net utilities should be maximized at values higher than the corresponding contracted rates.
- Later, when underusers wish to increase their rates because they have more data to send, they should have the incentive to do so and overusers should be encouraged to back down. Does their exist a pricing scheme such that this happens? Again, this means that the users' net utility values should be maximized at the appropriate points.

In [11], [12] and [13], the authors consider priority queuing to provide differentiated services to a mix of elastic and real-time traffic. Users choose the priority class to which their traffic belongs. Higher priority traffic experiences better service but its price is higher. Game-theoretic analysis is used to investigate whether a system equilibrium exists. [11] also considers how the network operator can set prices such that revenue is maximum at equilibrium. In our work, we do not have multiple classes and we do not consider priority queuing. There is only one traffic class, carrying elastic traffic.

II. MODEL

A. Utility function

We consider a single link which is shared among N users, N is a given and fixed integer. Capacity of the link is C bits/sec. We use a fluid model for traffic. User i generates fluid at rate λ_i. We emphasize that this is variable, because the user may not be able to generate traffic at a constant rate throughout.

User i has a contract to send traffic at rate γ_i, and the price charged by the SP is π_i; this is the *total price*, not the price per unit flow. When $\lambda_i > \gamma_i$, we call $(\lambda_i - \gamma_i)$ the "excess rate." Throughout this paper, we assume that the sum of the contracted rates equals the link capacity, *i.e.*, $\sum_{i=1}^{N} \gamma_i = C$.

If sum of contracted rates of all users happen to be less than capacity, the SP can either distribute the left over bandwidth equally among all users or keep it idle in hope of getting a future client. In the later case, we assume that there exists a virtual channel of capacity C' which is the sum of contracted rates of all existing users. The results and analysis of the paper will go through by simply replacing C by C'.

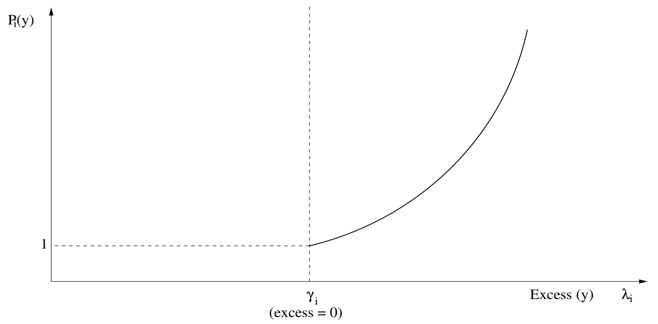

Fig. 1. An example congestion-penalty function. The domain is $\lambda_i \in [\gamma_i, \infty)$, *i.e.,* the domain corresponds to excess ≥ 0.

The *utility* of user i is a concave strictly increasing function of the rate of user i traffic *actually carried* by the SP. We assume that if a fraction β of the aggregate offered traffic is dropped, then a fraction β of *each* user's traffic is dropped. This means that the dropped traffic is split equally among all users. Then the utility for user i is $U_i(\lambda_i(1-\beta))$.

B. Multiplicative congestion-penalty function

The pricing scheme is characterized by the following features. *Underusers* are charged less than their contracted price. This is motivated by the goal of usage-based pricing. If a user is sending at a rate which is a fraction f of her contracted value, the charge is correspondingly a fraction f of her contracted charge. However, *overusers* are charged differently, depending upon whether the link is congested or not.

When the link is not congested, overusers are charged their contracted prices. The rationale for this is that as long as there is no congestion, users should be permitted to go above their contracted rates at no extra cost.

However, when the link is congested, overusers are charged heavily, because the overusers are themselves responsible for congestion. If overuser i is sending at a rate $\lambda_i > \gamma_i$, price charged is $\pi_i P_i(\lambda_i - \gamma_i)$, where $P_i(.)$ is a multiplicative congestion-penalty function, and it takes the excess rate $(\lambda_i - \gamma_i)$ as its argument. $P_i(\lambda_i - \gamma_i)$ is a convex increasing function of excess rate. Because the congestion-penalty function appears only when the link is congested and user i is an overuser, we set $P_i(0) = 1$. An example is shown in Figure 1.

In Figure 3, we give a schematic representation of the pricing scheme. We plot the price paid by user i versus her traffic rate λ_i. There are two curves: one for an uncongested link and the other for a congested link. As long as $\lambda_i < \gamma_i$, i is an underuser and is therefore charged less than π_i. When λ_i increases beyond γ_i, the price is maintained at π_i for the uncongested link; but for the congested link, the multiplicative congestion-penalty function appears and the price is $\pi_i P_i(\lambda_i - \gamma_i)$ which rises steeply.

C. Disutility function

Further, when user i is not able to send traffic at her contracted rate γ_i, *i.e.,* the rate λ_i is less than γ_i, we consider

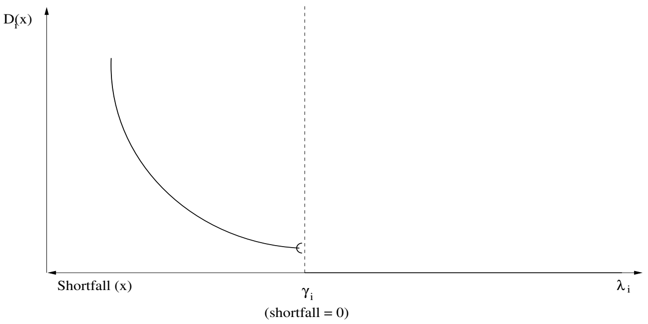

Fig. 2. An example disutility function. The function need not be continuous at $\lambda_i = \gamma_i$. Disutility is zero for $\lambda_i \geq \gamma_i$, and convex increasing in shortfall when the shortfall is strictly positive.

a "disutility" for user i. This measures the amount of dissatisfaction that user i suffers from at not being able to generate sufficient traffic. $(\gamma_i - \lambda_i)$ is referred to as the "shortfall" of user i, and the disutility function is a convex increasing function of shortfall. Further, the disutility function is defined to be zero when $\lambda_i \geq \gamma_i$. See Figure 2 for an example.

As mentioned in earlier sections, one of our objectives is to design a scheme in which underusers have the incentive to increase their rates of transmission when they have sufficient data to send. The disutility function is crucial in making this possible. For example, consider a user i who is below contracted rate and assume that other users are injecting traffic in such a manner that the network is already full. Now when user i wants to go to her contracted rate, net utility for her is $U_i(\lambda_i(1-\beta)) - (\frac{\lambda_i}{\gamma_i})\pi_i$. The utility is a function of the carried traffic. In this case, utility of the carried traffic is less than the utility of injected traffic. But, the price charged is proportional to the injected traffic. To let user i to increase flow, utility function should be "strong" enough to continuously give positive net utilities to the user till she reaches her contracted rate. A disutility function is important at this stage.

Furthermore, it can be shown mathematically that if we do not include a disutility function, only a restricted set of utility functions possessing certain specific mathematical structures can ensure fair and efficient operation of the network.

D. Net utility function

The "net utility" of user i is defined to be her utility minus disutility (which may be zero) minus the price paid to the SP.

III. Analysis

In this section, we explore how the prices π_i, $1 \leq i \leq N$, can be set so that all the desirable things happen. Intuitively, we can see that if these prices are too low they won't be able to restrict users from shooting above their contracted rates when there is congestion in the link. Also, prices cannot be exorbitantly high because then users won't have incentives to increase flows even to their contracted rates. Clearly, there is an upper and a lower bound on these prices. We would like to compute these bounds through the following discussions. We focus on user i and assume that the transmission rates λ_j of the other users $j, 1 \leq j \leq N, j \neq i$, are given and fixed.

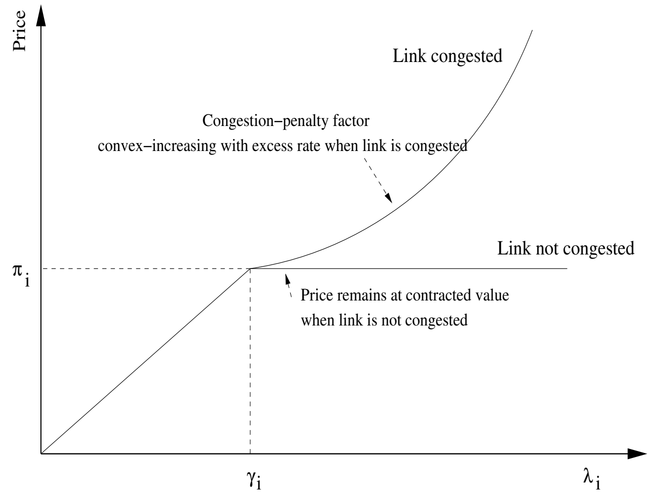

Fig. 3. Schematic showing how the price charged by the SP changes, depending on both the rate at which a user sends traffic and the state of congestion of the link.

Due to space constraints, we provide only a proof-outline for one of the following results in the Appendix. In all cases, the proofs are based on the definition of net utility for that case, elementary calculus and simple algebra. Throughout the rest of the paper we denote the aggregate sum of rates of all users except user i by θ_i, *i.e.*, $\theta_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \lambda_j$.

A. Other users below aggregate contracted rate

In this section, we consider a system where $\theta_i < (C - \gamma_i)$. In such a system, user i should have the incentive to transmit at a rate $(C - \theta_i)$, so that the link is fully utilized, but without causing congestion. We wish to find out how to set the price π_i such that this goal is achieved. The net utility function for user i is denoted as NU_i. Following are the three different situations which can arise depending upon the flow rate of user i. *(a)* When $\lambda_i < \gamma_i$, $NU_i = U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - (\frac{\lambda_i}{\gamma_i})\pi_i$, *(b)* when $\gamma_i \leq \lambda_i \leq (C - \theta_i)$, $NU_i = U_i(\lambda_i) - \pi_i$ and *(c)* when $\lambda_i > (C - \theta_i)$, $NU_i = U_i(\lambda_i(1-\beta)) - \pi_i P_i(\lambda_i - \gamma_i)$.

If $\frac{dNU_i}{d\lambda_i} > 0$, then i has incentive to increase her rate beyond λ_i. Corresponding conclusions apply when the derivative at λ_i is negative or zero.

Lemma 3.1: When the total traffic from users $j \in \{1, 2, ..., N\}$, $j \neq i$, is less than their aggregate contracted rate, NU_i is maximized at the point $\lambda_i^* = (C - \theta_i)$ if and only if the price π_i satisfies

$$\left(\frac{\theta_i}{C}\right)\frac{U_i'(C-\theta_i)}{P_i'(C-\theta_i-\gamma_i)} \leq \pi_i \leq \gamma_i(U_i'(\gamma_i) + \lim_{x\downarrow 0} D_i'(x)) \quad (1)$$

Further, NU_i is an increasing function of λ_i to the left of λ_i^* and a decreasing function of λ_i to the right of λ_i^*.
Here, $\lim_{x\downarrow 0} D_i'(x)$ indicates the right-hand limit of $D_i'(x)$ as x goes to zero while remaining positive always. This is necessitated by the definition of the disutility function $D(x)$ with x denoting the shortfall; as Figure 2 showed, $D(x)$ need not be continuous at $x = 0$.

B. Other users at aggregate contracted rate

Here we have $\theta_i = (C - \gamma_i)$. User i needs to transmit at a rate γ_i so that the link is fully utilized and yet there is no congestion. The net utility function for the different situations in this case are defined as follows. *(a)* When $\lambda_i < \gamma_i$, $NU_i = U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \big(\frac{\lambda_i}{\gamma_i}\big)\pi_i$ and *(b)* when $\lambda_i \geq \gamma_i$, $NU_i = U_i(\lambda_i(1-\beta)) - \pi_i P_i(\lambda_i - \gamma_i)$.

Lemma 3.2: When the total traffic from users $j \in \{1, 2, ..., N\}$, $j \neq i$, is equal to their aggregate contracted rate, NU_i is maximum at the point $\lambda_i^* = \gamma_i$ if and only if price π_i satisfies

$$\left(\frac{C-\gamma_i}{C}\right)\frac{U_i'(\gamma_i)}{\lim_{x\downarrow 0} P_i'(x)} \leq \pi_i \leq \gamma_i(U_i'(\gamma_i) + \lim_{x\downarrow 0} D_i'(x)) \quad (2)$$

Further, NU_i is an increasing function of λ_i to the left of λ_i^* and a decreasing function of λ_i to the right of λ_i^*.

We note that the interval in Equation 2 is a subset of that in Equation 1 because the lower bounds are ordered. So, a choice of π_i satisfying Equation 2 will always satisfy Equation 1.

C. Other users above aggregate contracted rate

Because other users are sending an aggregate amount of traffic that exceeds the sum of their contracted rates, we have $\theta_i > (C - \gamma_i)$. In this case, it is desirable that, notwithstanding the congested link, user i be able to increase her rate to her "rightful" share γ_i while other users back down. We can write the expressions for net utility in different situations as: *(a)*when $\lambda_i \leq (C - \theta_i)$, $NU_i = U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \big(\frac{\lambda_i}{\gamma_i}\big)\pi_i$, *(b)* when $(C - \theta_i) < \lambda_i < \gamma_i$, $NU_i = U_i(\lambda_i(1-\beta)) - D_i(\gamma_i - \lambda_i) - \big(\frac{\lambda_i}{\gamma_i}\big)\pi_i$ and *(c)* when $\lambda_i \geq \gamma_i$, $NU_i = U_i(\lambda_i(1-\beta)) - \pi_i P_i(\lambda_i - \gamma_i)$.

Lemma 3.3: When the total traffic from users $j \in \{1, 2, ..., N\}$, $j \neq i$, is above their aggregate contracted rate, NU_i is maximum at the point $\lambda_i^* = \gamma_i$ if and only if price π_i satisfies

$$\frac{\eta_i U_i'(\mu_i\gamma_i)}{\lim_{x\downarrow 0} P_i'(x)} \leq \pi_i \leq \gamma_i\left(\eta_i U_i'(\mu_i\gamma_i) + \lim_{x\downarrow 0} D_i'(x)\right) \quad (3)$$

where $\mu_i = \big(\frac{C}{\gamma_i+\theta_i}\big)$ and $\eta_i = \big(\frac{\mu_i\theta_i}{\gamma_i+\theta_i}\big)$. Further, NU_i is an increasing function of λ_i to the left of λ_i^* and a decreasing function of λ_i to the right of λ_i^*.

Having obtained the three ranges of π_i, we would like to know what the intersection of the three ranges looks like. The significance of this is that it may be possible to choose a value of π_i that will lead to desirable behavior irrespective of whether the aggregate traffic from all users except i is below, at or above the corresponding aggregate contracted rate.

Lemma 3.4: $\pi_i = \frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$ lies within all the three intervals if γ_i is such that

$$\gamma_i \geq \frac{U_i'(0)}{(\lim_{x\downarrow 0} D_i'(x))(\lim_{x\downarrow 0} P_i'(x))} \quad (4)$$

Proof. The left edge in Equation 1 is smaller than the left edge in Equation 2. Also, right edges in the two cases are same. So, it is enough to ensure that the second case is a valid interval. We know that utility is a concave increasing function and penalty is a convex increasing function. So

$$\Big(\frac{C-\gamma_i}{C}\Big)\frac{U_i'(\gamma_i)}{\lim_{x\downarrow 0} P_i'(x)} \leq \frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$$

If γ_i is such that it satisfies Equation 4,

$$\gamma_i(U_i'(\gamma_i) + \lim_{x\downarrow 0} D_i'(x)) \geq \frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$$

1 and 2 are non-void intervals and $\frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$ lies within both the intervals. Also, in 3, $\mu_i < 1$ and $\big(\frac{\theta_i}{\gamma_i+\theta_i}\big) \leq 1$ therefore making the lower bound less than $\frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$. If γ_i satisfies Equation 4, the upper bound will definitely exceed $\frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$ and we can choose $\pi_i = \frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$.

□

Lemma 3.5: If γ_i is such that

$$\gamma_i \geq \frac{\max_l U_l'(0)}{\Big(\min_l \lim_{x\downarrow 0} P_l'(x)\Big)\Big(\min_l \lim_{x\downarrow 0} D_l'(x)\Big)}$$
$$\forall i \in \{1, 2, \ldots, N\} \quad (5)$$

where the maximization and minimization are done over $l = \{1, 2, ..., N\}$, one single value of price for *all* users irrespective of congestion state of the link, is sufficient for the desirable behavior. That value of price is

$$\pi_i = \frac{\max_l U_l'(0)}{\min_l \lim_{x\downarrow 0} P_l'(x)}, \qquad \forall i \in \{1, 2, \ldots, N\} \quad (6)$$

Proof. We have already shown that for all users $i = \{1, 2, ..., N\}$, the lower bound of π_i can attain a maximum value of $\frac{U_i'(0)}{\lim_{x\downarrow 0} P_i'(x)}$. So if π_i is chosen according to Equation 6, for all users, π_i will exceed their respective lower limits.

If γ_i is such that it satisfies Equation 5, then

$$\begin{aligned} \gamma_i \lim_{x\downarrow 0} D_i'(x) &\geq \frac{\max_l U_l'(0) \lim_{x\downarrow 0} D_i'(x)}{\min_l \lim_{x\downarrow 0} P_l'(x) \min_l \lim_{x\downarrow 0} D_l'(x)} \\ &\geq \frac{\max_l U_l'(0)}{\min_l \lim_{x\downarrow 0} P_l'(x)} \end{aligned}$$

Therefore, upper bound on π_i for all $i = \{1, 2, ..., N\}$ is bigger than $\frac{\max_l U_l'(0)}{\min_l \lim_{x\downarrow 0} P_l'(x)}$.

□

IV. Experimental Results

In this section we provide experimental results of a simulation done in MATLAB. In this experiment 20 users access a link of capacity 27 units. We choose the following utility, penalty and disutility functions for all users. *(a)* $U(\lambda_i) = log(1 + \lambda_i)$, *(b)* $P(\lambda_i - \gamma_i) = 1 + (\lambda_i - \gamma_i)^2 + 2(\lambda_i - \gamma_i)$ and *(c)* $D(\gamma_i - \lambda_i) = (\gamma_i - \lambda_i)^2 + 2(\gamma_i - \lambda_i)$. According to Lemma 3.5, for all users contracted rates should be at least 0.25 and contracted price should be 0.5.

Accordingly, we divide these 20 users into 4 groups. Users 1, 2, 3 and 4 belong to the first group and they have a contracted rate of 0.5 units each. So the contracted rate vector for users in group 1 is γ_{g1} = [0.5 0.5 0.5 0.5]. Users 5, 6, 7, 8 and 9 belong to second group and have a contracted rate of 1 unit each. So γ_{g2} = [1 1 1 1 1] where the first element of the vector is the contracted rate of user 5 and so on. Group 3

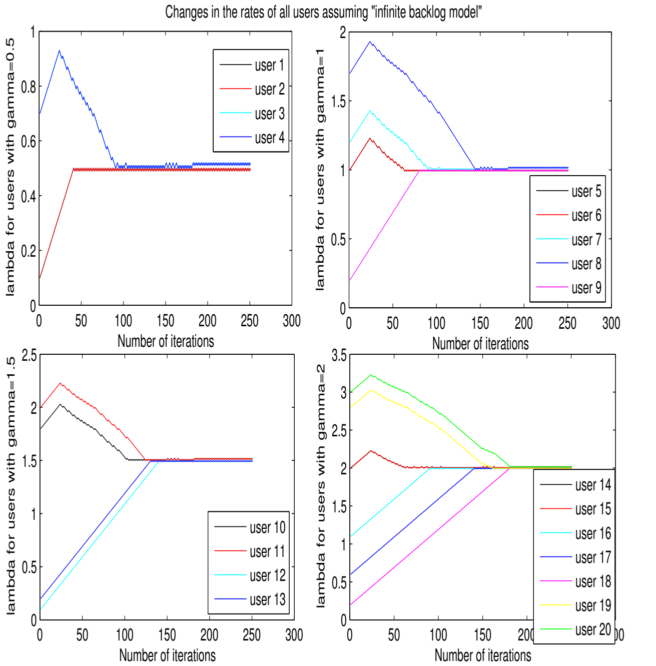

Fig. 4. Just before the start, λ =[λ_{g1} λ_{g2} λ_{g3} λ_{g4}]=[0.1 0.1 0.7 0.7 1.0 1.0 1.2 1.7 0.2 1.8 2.0 0.1 0.2 2.0 2.0 1.1 0.6 0.2 2.8 4.0]. At the beginning, all users have sufficient data to sustain any rate. The pricing scheme ensures that rates of all the users converge to their respective contracted values.

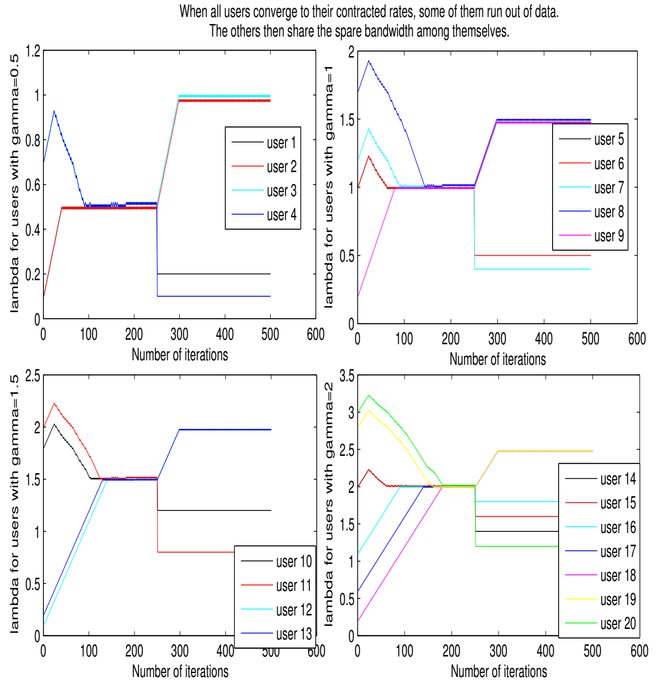

Fig. 5. Just before time 250, all users are at their contracted values. At time 250, users 1, 4, 6, 7, 10, 11, 14, 15, 16 and 20 are limited to 0.2, 0.1, 0.5, 0.4, 1.2, 0.8, 1.4, 1.6, 1.8 and 1.2 respectively. The pricing scheme ensures that users 2, 3, 5, 8, 9, 12, 13, 17, 18 and 19 go beyond their contracted rates and utilize the link fully.

consists of users 10, 11, 12 and 13 with γ_{g3} = [1.5 1.5 1.5 1.5]. And users 14, 15, 16, 17, 18, 19 and 20 form the fourth group having the contracted rate vector as γ_{g4} = [2 2 2 2 2 2 2]. The contracted rates add up to the capacity of the link. We choose an intial rate vector in which some users are underusers and some are overusers. For this experiment, we select the initial rates in the following manner. For the users in group 1 we choose rate vector to be λ_{g1} = [0.1 0.1 0.7 0.7]. So users 1 and 2 are underusers and users 3 and 4 are overusers. We set the rate vector for group 2 as λ_{g2} = [1.0 1.0 1.2 1.7 0.2]. Here users 7 and 8 are overusers and user 9 is underuser. Similarly, in case of groups 3 and 4, λ_{g3} = [1.8 2.0 0.1 0.2] and λ_{g4} = [2.0 2.0 1.1 0.6 0.2 2.8 4.0].

At this instant there is no data restriction and all users can sustain any rate. Each user explores the neighbourhood of her present rate to see where her net utility can be improved. We keep the step size as 0.01. We would expect underusers to go up and overusers to come down to their respective contracted rates. Figure 4 shows the experimental results for this situation. At the end of 250^{th} iteration, all the users have converged to their respective contracted rates.

Now we introduce data limitation on some users. Specifically, in group 1 we assume that users 1 and 4 run out of data and can sustain rates of only 0.2 and 0.1 units respectively. Therefore, at 251^{th} iteration, λ_{g1} = [0.2 0.5 0.5 0.1]. In group 2, we assume users 6 and 7 are stuck at 0.5 and 0.4 respectively therefore making λ_{g2} = [1.0 0.5 0.4 1.0 1.0]. Let us say users 10 and 11 in group 3 are data limited to 1.2 and 0.8 which gives λ_{g3} = [1.2 0.8 1.5 1.5]. Users 14, 15, 16 and 20 in group 4 are stuck at 1.4, 1.6, 1.8 and 1.2 respectively. So λ_{g4} = [1.4 1.6 1.8 2.0 2.0 2.0 1.2]. At this point, we would want users who are not data limited to utilize the spare room created by the underusers. We observe that price of 0.5 units for all users is sufficient for this to happen. Figure 5 shows the simulation results. At 500^{th} iteration, the flow rates converge to λ_{g1} = [0.200 0.975 0.995 0.100], λ_{g2} = [1.475 0.500 0.400 1.495 1.475], λ_{g3} = [1.200 0.800 1.975 1.975] and λ_{g4} = [1.400 1.600 1.800 2.475 2.475 2.475 1.200]. Although some users are underusers, cummulatively all users are able to use up 26.99 units of bandwidth out of the full capacity of 27 units.

At 501^{th} iteration, we assume that the underusers are back with data. Figure 6 shows that overusers back down and make room for underusers. By the end of 600^{th} iteration, all users again converge back to their respective contracted rates.

V. Stability

When no user is constrained by limited amounts of available data, what is the rate vector that the collection of users converges to? When prices are set according to Lemma 3.5, it can be seen that $(\gamma_1, \gamma_2, \ldots, \gamma_N)^t$ is a Nash equilibrium. Considering any $i \in \{1, 2, \ldots, N\}$, we find that aggregate rate from other users is equal to the sum of their contracted rates, and therefore, Lemma 3.2 applies. This means that user i has no incentive to change her rate from γ_i, and this conclusion applies to all the users.

Moreover, it can be seen that $(\gamma_1, \gamma_2, \ldots, \gamma_N)^t$ is the unique Nash equilibrium in this problem and system stability is assured. We consider a $\overline{\lambda}$ with $\sum_{j=1}^{N} \lambda_j = C$, but with one i for which $\lambda_i > \gamma_i$. Then, there must be a k for which $\lambda_k < \gamma_k$. With respect to this k, $\sum_{\substack{j=1 \\ j \neq k}}^{N} \lambda_j > \sum_{\substack{j=1 \\ j \neq k}}^{N} \gamma_j$, and the third case above applies. Therefore, by Lemma 3.3, k would like

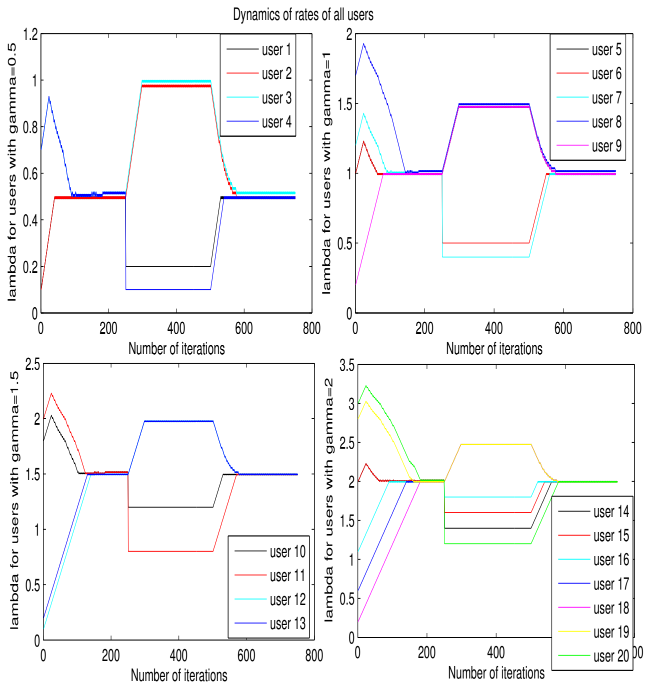

Fig. 6. Continuing the experiment in Figures 4 and 5. At time 500, underusers 1, 4, 6, 7, 10, 11, 14, 15, 16 and 20 are again ready to send data at high rates. The pricing scheme ensures that underusers reclaim their bandwidth shares while overusers 2, 3, 5, 8, 9, 12, 13, 17, 18 and 19 back down.

to increase her rate to γ_k. Hence $\overline{\lambda}$ is not a Nash equilibrium. In other cases as well ($\sum_{j=1}^{N} \lambda_j > C$ and $\sum_{j=1}^{N} \lambda_j < C$) we can argue similarly.

VI. Conclusion

We considered sources that could occasionally be constrained by limited amounts of available data. Further, each user has a contract with the service provider, specifying the rate at which she can send traffic into the network. We were interested in a congestion-dependent as well as user-dependent pricing scheme that would ensure fair and efficient sharing of the single link shared by the sources.

We introduced the idea of disutility for underusers and noted that the disutility term encourages underusers to increase their rates whenever they have sufficient data. We presented simple necessary and sufficient conditions for setting prices such that fair and efficient sharing of the link is possible and observed that one value of price for all users can achieve this. The SP is not required to store different prices for different users and keep changing them dynamically depending upon the congestion state of the link. A simple experiment in MATLAB demonstrated the utility of our approach.

We recognize the following limitations of our work. Firstly, we considered the resource to be only a single link; in general, of course, the resource is a full network. This is the topic of ongoing work. Secondly, we have not explicitly considered the problem of revenue maximization for the service provider. While our goals of fair and efficient sharing of the link are natural and would, indirectly, lead to high revenue for the provider, we would like to consider the problem of explicit revenue maximization as well.

References

[1] F. Kelly, "Charging and rate control for elastic traffic (corrected version)," *European Transaction on Telecommunication*, vol. 8, no. 1, pp. 33–37, Jan 1997.

[2] F. Kelly, A. K. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, Mar 1998.

[3] R. J. La and V. Anantharam, "Utility-based rate control in the internet for elastic traffic," *IEEE Transactions on Networking*, vol. 10, no. 2, pp. 272–286, Apr 2002.

[4] 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, Oct 2000.

[5] J. Shu and P. Varaiya, "Pricing network services," *Proceedings of the IEEE Infocom*, Mar 2003.

[6] T. Alpcan and T. Basar, "A utility-based congestion control scheme for internet-style networks with delay," *Proceedings of IEEE InfoCom*, Apr 2003.

[7] S. Kunniyar and R. Srikant, "End-to-end congestion control schemes: Utility functions, random losses and ecn marks," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 689–702, Oct 2003.

[8] A. A. Lazar and N. Semret, "Design and analysis of the progressive second bid auction for network bandwidth sharing," *Telecommunication Systems – Special Issue on Network Economics*, 1999.

[9] A. Delenda, "Mécanismes dèncheres pour le partage de ressources télécom," France Telecom – R & D, Tech. Rep. Tech Rep 7831, 2002.

[10] P. Maillé and B. Tuffin, "Multi-bid auctions for bandwidth allocation in communication networks," in *Proceedings of IEEE Infocom*, April 2004.

[11] M. Mandjes, "Pricing strategies under heterogeneous service requirements," *Proceedings of the IEEE Infocom*, Mar 2003.

[12] P. Marbach, "Analysis of a static pricing scheme for priority services," *IEEE/ACM Transactions on Networking*, vol. 12, no. 2, pp. 312–325, Apr 2004.

[13] ——, "Priority service and max-min fairness," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 733–746, Oct 2003.

Appendix

Proof of Lemma 3.1

When $\lambda_i < \gamma_i$, the net utility (NU_i) is $U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \left(\frac{\lambda_i}{\gamma_i}\right)\pi_i$. i increases her rate if and only if $\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i} \geq 0$. That is, $U_i'(\lambda_i) + D_i'(\gamma_i - \lambda_i) - \left(\frac{\pi_i}{\gamma_i}\right) \geq 0$. To let NU_i increase till λ_i reaches γ_i, it is sufficient to have $\pi_i \leq \gamma_i(U_i'(\gamma_i) + \lim_{x\downarrow 0} D_i'(x))$. This condition is also necessary because if $\pi_i = \gamma_i(U_i'(\gamma_i) + \lim_{x\downarrow 0} D_i'(x)) + \epsilon$, where $\epsilon > 0$, then $\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i}\Big|_{\lambda_i \uparrow \gamma_i} = -\frac{\epsilon}{\gamma_i}$. User i looses incentive to increase rate even before reaching contracted rate.

When $\lambda_i \in [\gamma_i, (C - \theta_i)]$, NU_i is simply $U_i(\lambda_i) - \pi_i$ which is an increasing function of λ_i. This in turn gives user i every reason to increase λ_i in this range.

When $\lambda_i > (C - \theta_i)$, link gets congested. As long as $\pi_i \geq 0$ and $P_i(C - \theta_i - \gamma_i) \geq 1$, NU_i just above $(C - \theta_i)$ is less than NU_i at $(C - \theta_i)$. To ensure that $\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i} \leq 0$ when $\lambda_i > (C - \theta_i)$, let $\lambda_i = C - \theta_i + \delta$, for a $\delta > 0$. By definition, $NU_i = U_i((C - \theta_i + \delta)(1 - \beta)) - \pi_i P_i(C - \theta_i + \delta - \gamma_i)$ where $\beta = \frac{(\lambda_i + \theta_i - C)}{(\lambda_i + \theta_i)} = \frac{\delta}{(C + \delta)}$. Differentiating this net utility with respect to δ and choosing a π_i which keeps the derivative non-positive, we get $\pi_i \geq \frac{C\theta_i}{(C+\delta)^2} U_i'\left(\frac{(C-\theta_i+\delta)C}{(C+\delta)}\right) \frac{1}{P_i'(C-\theta_i-\gamma_i+\delta)}$. Because we want the lower bound to hold for *every* $\delta > 0$, we now take supremum of the lower bound over all $\delta > 0$ which yields $\pi_i \geq \left(\frac{\theta_i}{C}\right) \frac{U_i'(C-\theta_i)}{P_i'(C-\theta_i-\gamma_i)}$. To prove the necessity of the lower bound let $\epsilon > 0$ be a given small number and we take $\left(\left(\frac{\theta_i}{C}\right) \frac{U_i'(C-\theta_i)}{P_i'(C-\theta_i-\gamma_i)} - \epsilon\right)$ as the lower bound, then simple algebra shows that there exists a $\delta > 0$ such that $\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i} > 0$. This contradicts our requirement and concludes the proof.