Pricing Network Resources: A New Perspective

Joy Kuri
Centre for Electronics Design and Technology
Indian Institute of Science
Bangalore, India
kuri@cedt.iisc.ernet.in

Sharmili Roy
Centre for Electronics Design and Technology
Indian Institute of Science
Bangalore, India
sharmili@cedt.iisc.ernet.in

ABSTRACT

We consider I users sending elastic traffic into a Service Provider's (SP's) network. Each user has a contract to transmit data at a nominal rate. However, a user is free to transmit at a higher rate if she wishes.

For each user, we assume a realistic traffic model. A user alternates between two phases over time. In the *limited data* phase, a user's transmission rate is capped at some value below her contracted rate. This happens because the user does not have enough data to send. In the *unlimited data* phase, a user has enough data to sustain any transmission rate.

At time t_1, a (possibly empty) set S_1 of users is in the limited data phase. This leaves unutilized resources in the network that ought to be exploited by others. At time $t_2 > t_1$, users in S_1 are back in the unlimited data phase. Now it should be possible for these users to reclaim their contracted rates at the expense of "overusers."

We show that under certain conditions, a very simple pricing scheme can ensure fair and efficient operation in the above sense. In our scheme, a SP needs to maintain just one single price for all the users.

Categories and Subject Descriptors

C.2.5 [**local and Wide-Area Networks**]: Access Schemes

General Terms

Design, Performance, Theory

Keywords

Pricing, Congestion, Fairness

1. 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 ([3] [4] [6] [12] [13], [1] [5]), 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 a 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 [3], 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.

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 ([7], [2]). Accordingly, we assume in this paper that users' utility functions are known to the central authority (the SP) 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, [8] 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 [8], our objective in this paper is the former, *viz.,* fair and efficient sharing of a network; 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 SP that specify the rates at which they can send traffic into the network. We are interested in devising 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. However, even though our framework allows different prices for different users, our analysis shows that under some conditions that are easily satisfied, 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.

We are interested in ensuring that network operation is characterized 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.

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

In [9], [11] and [10], 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. [9] also considers how the network operator can set prices such that revenue is maximized 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.

2. MODEL

2.1 Utility function

We consider a network which is shared among I users, where I is a given and fixed integer. We use a fluid model for traffic. User i injects fluid at rate λ_i into the network. We emphasize that this is a 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." The *utility* of user i is a concave strictly increasing function of the rate of user i traffic *actually carried* from the source to destination node. If λ_i is the flow i injects at source node and some part of i's traffic is dropped, then $\mu_i \leq \lambda_i$ is the amount carried to the destination node. The utility for user i is $U_i(\mu_i[\lambda_i])$. Since carried traffic (μ_i) depends upon the injected traffic (λ_i), we denote μ_i by $\mu_i[\lambda_i]$ to emphasize this dependence.

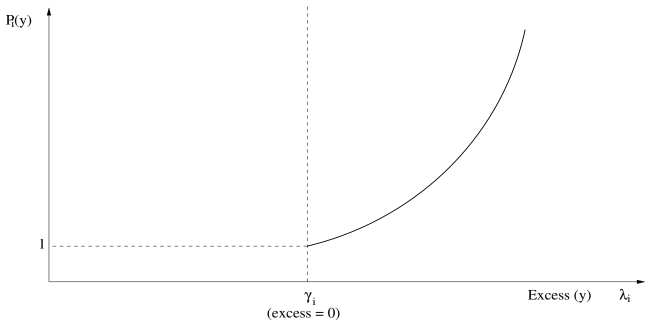

Figure 1: An example congestion-penalty function. The domain is $\lambda_i \in [\gamma_i, \infty)$, *i.e.,* the domain corresponds to excess ≥ 0.

2.2 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 network is congested or not.

When the network 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 network is congested, overusers are charged heavily, because the overusers are themselves responsible for the congestion. If overuser i is sending at a rate $\lambda_i > \gamma_i$, then the 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 network is congested and user i is an overuser, we set $P_i(0) = 1$. An example is shown in Figure 1.

In Figure 2, 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 network and the other for a congested network. 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 network; but for the congested network, the multiplicative congestion-penalty function appears, the price is $\pi_i P_i(\lambda_i - \gamma_i)$ and the price paid rises steeply.

2.3 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 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 de-

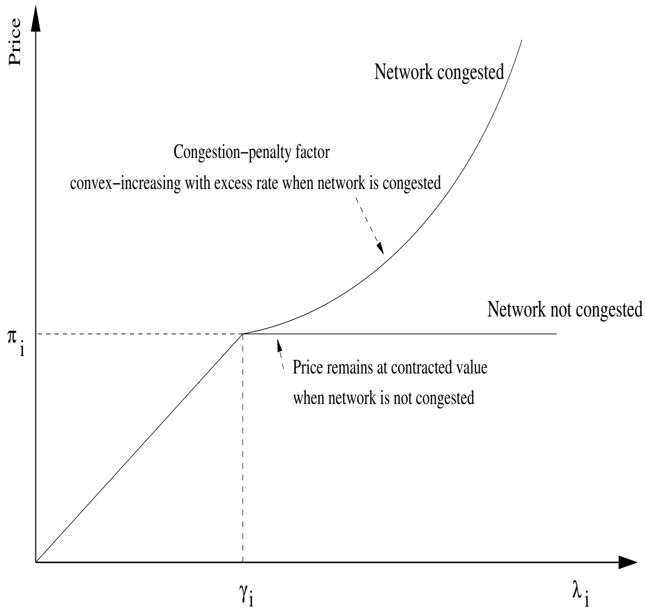

Figure 2: Schematic showing how the price charged by the SP changes, depending on both the rate at which a user sends traffic and the state of congestion of the network.

fined to be zero when $\lambda_i \geq \gamma_i$. See Figure 3 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 the 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, the net utility for her is $U_i(\mu_i[\lambda_i]) - \left(\frac{\lambda_i}{\gamma_i}\right)\pi_i$. 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 increase flow, utility function should be "strong" enough to continuously give positive net utilities to the user

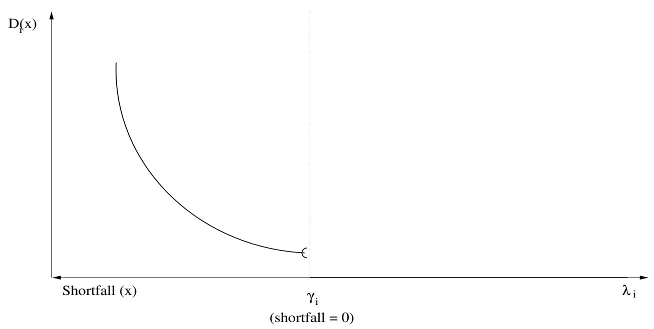

Figure 3: An example disutility function. The function need not be continuous at $\lambda_i = \gamma_i$. Disutility is zero for $\lambda_i \geq \gamma_i$, and convex increasing in shortfall when the shortfall is strictly positive.

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, the prices for all users do not remain equal and they need to be continuously modified as the network evolves. Also, these new prices have to conveyed back to the users as and when they get altered failing which the network may not behave in the way we want it to.

2.4 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.

2.5 Fair and Efficient Network Sharing

Our intention is to understand how the prices π_i, $1 \leq i \leq I$, can be set, so that the system behavior is "desirable" in the following sense. Let $\mathcal{L}$ and $\mathcal{H}$ denote the set of underusers and overusers, respectively.

When the resource is not congested because users in $\mathcal{L}$ are below their contracted rates, other users in the complement $\mathcal{L}^c$ are not prevented from going above their contracted rates. This is desirable so that the resource is fully utilized.

Suppose that the resource is fully utilized and, further, that there is no congestion. Now if users in $\mathcal{L}$ wish to increase their rates and go up to their contracted values, they should have the incentive to do that; at the same time, users in $\mathcal{H}$ should have the incentive to reduce their rates. This is desirable so that users in $\mathcal{L}$ can reclaim their share of the network resources from users in $\mathcal{H}$ when they wish to do so.

3. ANALYSIS

In this section, we explore how the prices π_i, $1 \leq i \leq I$, can be set so that desirable behavior is enforced. Intuitively, if these prices are too low, users cannot be prevented from transmitting above their contracted rates even if there is congestion in the network. Also, prices cannot be too high because then users will not have incentive 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. We focus on user i and assume that the transmission rates λ_j of the other users $j, 1 \leq j \leq I, j \neq i$, are given and fixed.

Let us consider a network which has N nodes and L links. The network is characterized by the node link incidence matrix A and a set of flow vectors, say $y^{(i)}$'s, each associated with a user i. An element $y^{(i)}(l)$, $l = \{1, 2, \ldots, L\}$ is the amount of user i's traffic carried on link l.

Suppose that there is a user i who injects λ_i amount of traffic at source node s_i. Let us call the destination node for this user as node z_i. There is a vector $d^{(i)}$ of size $N \times 1$ corresponding to each user i such that an element $d^{(i)}(n)$, $n = \{1, 2, \ldots, N\}$, gives amount of i's traffic dropped at node n. The flow conservation equation for user i is

$$Ay^{(i)} + d^{(i)} = v^{(i)} \tag{1}$$

where $v^{(i)}(s_i) = \lambda_i$, $v^{(i)}(z_i) = -\mu_i[\lambda_i]$ and all other elements are 0, *i.e.*, $v^{(i)} = (00\ldots\lambda_i 0\ldots - \mu_i[\lambda_i]\ 0\ldots)^t$. $\mu_i[.]$ denotes the function which gives the traffic delivered to the destination node. Since some traffic may be dropped, $\mu_i[\lambda_i] \leq \lambda_i$. Flow conservation (Equation 1) should hold for all users.

Also, the link capacity constraints are

$$\sum_{i=\{1,2,\ldots,I\}} y^{(i)}(l) \leq C_l, \qquad \forall\ l = \{1,2,\ldots,L\} \tag{2}$$

where C_l is the capacity of l^{th} link.

One question which may arise at this point is; given the injected traffic of all users (λ_i's), how do we know their flow vectors ($y^{(i)}$'s), vector of drops ($d^{(i)}$'s) and the carried traffic ($\mu_i[\lambda_i]$'s)? There can be multiple $y^{(i)}$'s, $d^{(i)}$'s and $\mu_i[\lambda_i]$'s which satisfy the flow conservation (Equation 1) and capacity constraint equations (Equation 2). To get a unique value of $y^{(i)}$'s, $d^{(i)}$'s and $\mu_i[\lambda_i]$'s, the network can perform some kind of an optimization subject to the flow conservation and capacity constraints. The optimization can be maximization of sum of utility functions of all users, maximization of the sum of carried traffic of all users, maximization of the net revenue earned by the SP to name a few.

Now suppose λ_j, $j \in \mathcal{I} \setminus \{i\}$, are given and fixed. We define a congestion threshold ct_i for user i as the maximum injected traffic λ_i till which i's traffic is not dropped in the network. From user i's perspective, three possibilities arise: *(a)* the network gets congested after the user is above her contracted rate, *i.e.*, $ct_i > \gamma_i$, *(b)* there is congestion in the network as soon as i tries to exceed contracted rate, *i.e.*, $ct_i = \gamma_i$ and *(c)* even before user reaches her contracted rate, the network becomes congested, *i.e.*, $ct_i < \gamma_i$. We consider these three cases one by one in the following paragraphs.

The proof-outlines of the following results are provided in the Appendix. In all cases, the proofs are based on the definition of net utility for that case, elementary calculus and simple algebra.

3.1 Congestion threshold above contracted rate

In this section, we consider a system where $ct_i > \gamma_i$. This means that even if user i transmits at her contracted rate, there will be some spare capacity left in the network. In such a situation, it is desirable to let user i transmit at ct_i, so that the network 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 cases which can arise depending upon the flow rate of user i.

- When $\lambda_i < \gamma_i$,

$$NU_i \;=\; U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i$$

- When $\gamma_i \leq \lambda_i \leq ct_i$,

$$NU_i \;=\; U_i(\lambda_i) - \pi_i$$

- When $\lambda_i > ct_i$,

$$NU_i \;=\; U_i(\mu_i[\lambda_i]) - \pi_i P_i(\lambda_i - \gamma_i)$$

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 $j \in \{1, 2, ..., I\}$, $j \neq i$, is such that even if user i transmits at γ_i, there is some spare capacity left in the network, NU_i is maximized at the point $\lambda_i^* = ct_i$ if and only if the price π_i satisfies*

$$\frac{U_i'(\mu_i[ct_i + \delta^*])\mu_i'[ct_i + \delta^*]}{P_i'(ct_i + \delta^* - \gamma_i)} \leq \pi_i \leq \gamma_i(U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x)) \tag{3}$$

where $\delta^* = \arg\max_{\delta>0}\big(\frac{U_i'(\mu_i[ct_i+\delta])\mu_i'[ct_i+\delta]}{P_i'(ct_i+\delta-\gamma_i)}\big)$. $\mu_i[ct_i+\delta^*]$ is user i's flow received at the destination node when $(ct_i+\delta^*)$ is the flow injected at the source node. 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, $\lim_{x\to 0+} D_i'(x)$ indicates the right-hand limit of $D_i'(x)$ as x goes to zero, *i.e.*, 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 3 showed, $D(x)$ need not be continuous at $x = 0$.

3.2 Congestion threshold equal to contracted rate

Here we have $ct_i = \gamma_i$. This means, as soon as user i tries to go above contracted rate, the network is over-full. Our objective here is to set the price such that user has incentive only to go up to her contracted rate. The net utility function for different situations in this case are defined as follows.

- When $\lambda_i < \gamma_i$,

$$NU_i \;=\; U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i$$

- When $\lambda_i \geq \gamma_i$,

$$NU_i \;=\; U_i(\mu_i[\lambda_i]) - \pi_i P_i(\lambda_i - \gamma_i)$$

LEMMA 3.2. *When the total traffic from users $j \in \{1, 2, ..., I\}$, $j \neq i$, is such that, when user i transmits at γ_i the network is full, NU_i is maximized at the point $\lambda_i^* = \gamma_i$ if and only if price π_i satisfies*

$$\frac{U_i'(\mu_i[\gamma_i + \delta^*])\mu_i'[\gamma_i + \delta^*]}{P_i'(\delta^*)} \leq \pi_i \leq \gamma_i(U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x)) \tag{4}$$

where $\delta^* = \arg\max_{\delta>0}\big(\frac{U_i'(\mu_i[\gamma_i+\delta])\mu_i'[\gamma_i+\delta]}{P_i'(\delta)}\big)$. $\mu_i[\gamma_i+\delta^*]$ is the amount of user i's flow received at the destination node when $(\gamma_i+\delta^*)$ is the flow injected at the source node. 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^*.

3.3 Congestion threshold below contracted rate

Because the congestion threshold is below the contracted rate, i does not even have enough room to go up to her contracted rate without causing congestion in the network. Nevertheless, it is desirable that the user increase her rate to the "rightful" share γ_i while other users back down. We can write the expressions for net utility in different cases as

- When $\lambda_i \leq ct_i$,

$$NU_i \;=\; U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i$$

- When $ct_i < \lambda_i < \gamma_i$,

$$NU_i \;=\; U_i(\mu_i[\lambda_i]) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i$$

- When $\lambda_i \geq \gamma_i$,

$$NU_i \quad = \quad U_i(\mu_i[\lambda_i]) - \pi_i P_i(\lambda_i - \gamma_i)$$

Lemma 3.3. *When the total traffic from users $j \in \{1, 2, .., I\}$, $j \neq i$, is such that even if user i transmits at a rate less than γ_i, the network is congested, NU_i is maximized at the point $\lambda_i^* = \gamma_i$ if and only if price π_i satisfies*

$$\frac{U_i'(\mu_i[\gamma_i + \delta^*])\mu_i'[\gamma_i + \delta^*]}{P_i'(\delta^*)} \leq \pi_i \leq \min(\gamma_i \eta_i, \gamma_i \zeta_i) \qquad (5)$$

where $\eta_i = U_i'(ct_i) + D_i'(\gamma_i - ct_i)$ and $\zeta_i = U_i'(\mu_i[ct_i + \delta^+])\mu_i'[ct_i + \delta^+] + D_i'(\gamma_i - ct_i - \delta^+)$. $\mu_i[\gamma_i + \delta^*]$ is amount of user i's flow received at the destination node when $(\gamma_i + \delta^*)$ is the flow injected at source node. δ^+ is defined as $\arg\min_{0 \leq \delta \leq (\gamma_i - ct_i)} \gamma_i \big(U_i'(\mu_i[ct_i + \delta])\mu_i'[ct_i + \delta] + D_i'(\gamma_i - ct_i - \delta)\big)$ and $\delta^* = \arg\max_{\delta > 0} \big(\frac{U_i'(\mu_i[\gamma_i + \delta])\mu_i'[\gamma_i + \delta]}{P_i'(\delta)}\big)$. Further, NU_i is an increasing function of λ_i to the left of λ_i^* and a decreasing funct in of λ_i to the right of λ_i^*.

Remark: The lower bounds in Equations 3, 4 and 5 are the ratio of user i's marginal utility and marginal penalty incurred at the carried traffic (just above λ_i^*) multiplied by the marginal carried traffic. If marginal utility or marginal carried traffic in going above λ_i^* is high and marginal penalty is low, then the user is already inclined to increase her rate. To prevent this from happening, a high price must be set. Hence, in this situation, the lower bound on π_i is large. On the other hand, if marginal penalty is high and marginal utility and marginal carried traffic are low, the user is already disinclined to increase her rate. Even a small price is sufficient in this case, and the lower bound on π_i is small.

Now let us focus our attention on the upper bounds. Treating $\left(\frac{\pi_i}{\gamma_i}\right)$ as the effective "price per unit bandwidth", we see that the upper bound says that: Price per unit bandwidth should be less than the sum of user i's marginal utility when transmitting at the contracted rate and the marginal disutility at a positive shortfall. If the unit price exceeds the sum above, there is no motivation for the user to increase the flow up to even the contracted value.

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 work (*i.e.,* lead to desirable behavior) irrespective of whether the congestion threshold is above, below or equal to contracted rate.

Lemma 3.4. *We can set $\pi_i = \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$ irrespective of the congestion state of the network, if we choose congestion penalty function such that*

$$\lim_{x \to 0+} P_i'(x) \geq \frac{U_i'(0)}{\gamma_i \lim_{x \to 0+} D_i'(x)} \qquad (6)$$

Proof. Let us consider Equation 3 first. Owing to the concave increasing nature of utility function and convex increasing nature of congestion penalty function,

$$\begin{aligned} U_i'(\mu_i[ct_i + \delta^*]) &\leq U_i'(0) \\ P_i'(ct_i + \delta^* - \gamma_i) &\geq \lim_{x \to 0+} P_i'(x) \end{aligned}$$

Since δ^* is a positive quantity,

$$\begin{aligned} \mu_i[ct_i + \delta^*] &\leq ct_i + \delta^* \\ \mu_i'[ct_i + \delta^*] &\leq 1 \end{aligned}$$

From these observations, we can infer that

$$\frac{U_i'(\mu_i[ct_i + \delta^*])\mu_i'[ct_i + \delta^*]}{P_i'(ct_i + \delta^* - \gamma_i)} \leq \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$$

Following exactly similar arguments, we can show that the lower bounds of Equations 4 and 5 are also never more than $\frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$.

Now, if we choose congestion penalty function according to Equation 6,

$$\begin{aligned} \gamma_i \lim_{x \to 0+} D_i'(x) &\geq \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)} \\ \gamma_i \Big(U_i'(\gamma_i) + \lim_{x \to 0+} D_i'(x)\Big) &\geq \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)} \end{aligned}$$

The consequence is, upper bounds in Equations 3 and 4 are always to the right of $\frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$. Now consider Equation 5. In the context of this equation, $\delta^+ \geq 0$. So,

$$D_i'(\gamma_i - ct_i) \geq D_i'(\gamma_i - ct_i - \delta^+) \geq \lim_{x \to 0+} D_i'(x)$$

If congestion penalty is in accord with Equation 6,

$$\begin{aligned} \gamma_i D_i'(\gamma_i - ct_i - \delta^+) &\geq \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)} \\ \gamma_i D_i'(\gamma_i - ct_i) &\geq \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)} \end{aligned}$$

So, even the upper bound of Equation 5 is never less than $\frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$. Thus, the intersection interval for each user i is non-empty, and $\pi_i = \frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$ lies within this intersection interval. □

Thus, if $\lim_{x \to 0+} P_i'(x)$ satisfies (6), it is possible to set π_i appropriately without knowing ct_i. This is clearly advantageous.

Thus, the SP can pick one price which lies within the intersection of the three ranges. Next we explore whether there is one price which lies within the intersection of the three ranges of *all* users sharing the network.

Lemma 3.5. *If we choose congestion penalty functions such that*

$$\lim_{x \to 0+} P_i'(x) \geq \frac{\max_k U_k'(0)}{\min_k \gamma_k \min_k \lim_{x \to 0+} D_k'(x)}, \forall i \in \{1, 2, \ldots, I\} \qquad (7)$$

where the maximization and minimization are done over $k = \{1, 2, ..., I\}$, one single value of price for all users, irrespective of the network behavior, is sufficient to enforce desirable system operation. That value of price is

$$\pi_i = \frac{\max_k U_k'(0)}{\min_k \lim_{x \to 0+} P_k'(x)}, \qquad \forall \ i \in \{1, 2, \ldots, I\} \qquad (8)$$

Proof. We have already shown that for all users $i = \{1, 2, ..., I\}$, the lower bound on π_i can attain a maximum of $\frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)}$ only. And,

$$\frac{U_i'(0)}{\lim_{x \to 0+} P_i'(x)} \leq \frac{\max_k U_k'(0)}{\min_k \lim_{x \to 0+} P_k'(x)}$$

So if π_i's are chosen according to Equation 8, for all users, π_i's will exceed their respective lower limits.

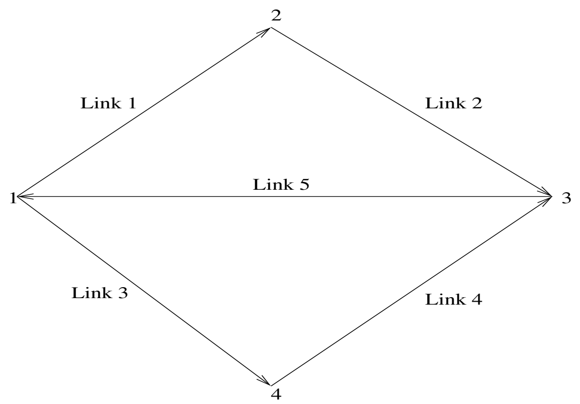

Figure 4: Network considered for the experiment.

Table 1: Source Node, Destination Node, Contracted rate, Utility, Disutility and Penalty functions of the four users

	s_i	z_i	γ_i	$U(x)$	$D(x)$	$P(x)$
User 1	1	3	1.0	x	e^x	$1+x^2+2x$
User 2	1	4	1.5	$log(1+x)$	x	e^x
User 3	3	2	2.0	x	x^2+2x	$1+x$
User 4	3	4	1.8	$1-e^{-x}$	$(1+x)^3$	e^x

If congestion penalty satisfies Equation 7, then

$$\min_k \gamma_k \min_k \lim_{x\to 0+} D'_k(x) \geq \frac{\max_k U'_k(0)}{\min_k \lim_{x\to 0+} P'_k(x)}$$

$$\gamma_i \lim_{x\to 0+} D'_i(x) \geq \frac{\max_k U'_k(0)}{\min_k \lim_{x\to 0+} P'_k(x)}$$

Therefore, the upper bound of π_i for all $i = \{1, 2, ..., I\}$ is bigger than $\frac{\max_k U'_k(0)}{\min_k \lim_{x\to 0+} P'_k(x)}$ thereby letting us choose π_i according to Equation 8. $\square$

The conclusion is that although all users are only interested in maximizing their net benefits, one price — the same for all users — compels them to behave in the system optimal manner also.

4. SIMULATION

In this section we present the results of a MATLAB simulation. In this experiment, 4 users share the network shown in Figure 4. We study how the rates of these users evolve when they alternate between limited and unlimited data phases.

Each user is associated with one flow. We use the words flow and user interchangeably. The source/destination nodes, contracted rates and utility, disutility and congestion penalty functions for these users are as tabulated in Table 1. The first observation is that all congestion penalty functions obey Equation 7. So, a contracted price of 1 unit (Equation 8) is applicable to all users.

The SP conveys the contracted price and penalty function to the users. Also, whether the network is congested or not

Table 2: Initial operating point of the network

	Link 1	Link 2	Link 3	Link 4	Link 5
User 1	1.2	1.2	0.8	0.8	0.0
User 2	0.0	0.0	1.0	0.0	0.0
User 3	1.8	0.0	0.0	0.0	1.8
User 4	0.0	0.0	1.5	0.0	1.5

comes as a feedback from the SP to the users.[1] Once the SP knows the set of underusers and users with unlimited data, he computes the operating point of the network in terms of the flow vectors $y^{(i)}$'s, the vector of drops $d^{(i)}$'s and the carried traffics $\mu_i[\lambda_i]$'s of all users. For this experiment we assume the optimization problem for the SP to be the following.

$$\max_{y^{(i)}, d^{(i)}, \mu_i[\lambda_i], \forall i\in\mathcal{I}} \quad \sum_{i\in\mathcal{I}} w_i \mu_i[\lambda_i] \tag{9}$$

$$\begin{aligned} \text{subject to} \quad Ay^{(i)} + d^{(i)} &= v^{(i)}, & \forall\ i \in \mathcal{I} \\ \sum_{i\in I} y^{(i)}(l) &\leq C_l, & \forall\ l \in \mathcal{L} \\ \mu_i[\lambda_i] &\leq \lambda_i, & \forall\ i \in \mathcal{I} \\ y^{(i)}, d^{(i)}, \mu_i[\lambda_i] &\geq 0, & \forall\ i \in \mathcal{I} \end{aligned}$$

where w_i is a weight/preference associated with user i. In our example all users have equal preference. The SP can find out $y^{(i)}$'s, $d^{(i)}$'s and $\mu_i[\lambda_i]$'s by solving this problem.

Now if at a later point of time, the set of underusers changes — because some underusers now have enough data or a new group of users are now data-limited — the SP reroutes the flows by re-solving the maximization problem in order to maximally utilize the network.

At the start of the simulation, we assume that the initial data rates are $\lambda = [2.0\ 1.0\ 1.8\ 1.5]$, *i.e.*, user 1's data rate is 2.0, user 2's data rate is 1.0 and likewise. So, to begin with user 1 is an overuser and users 2, 3 and 4 are underusers. The SP routes the initial flows as shown in Table 2. There are no drops in the network at this point.

Now suppose that all users have unlimited data to sustain any rate. At each time instant, they try to maximize their respective net utilities. At each iteration of the simulation, all users perturb their present data rates and compute new net utilities at these perturbed rates. If the net utility increases on increasing rate, users increase their flow rates in the next iteration; similarly, if the net utility increases on decreasing rate, users decrease rates in the next iteration. For this MATLAB experiment, we assume the step-size to be 0.01, equal for all, *i.e.*, at each iteration λ_i's increase/decrease by 0.01 units or stay put.

On running the simulation, we find that within 100 iterations, all users converge to their respective contracted rates in the process of maximizing their net utilities. This is because the links are provisioned such that the network has just enough capacity to accommodate all users at their respective contracted rates. Figure 5 shows the simulation results and routes taken by the flows are tabulated in Table 3.

[1]In this paper, we make the simplifying assumption that all feedback from the network to the users is available instantaneously. In practice, signalling is necessary.

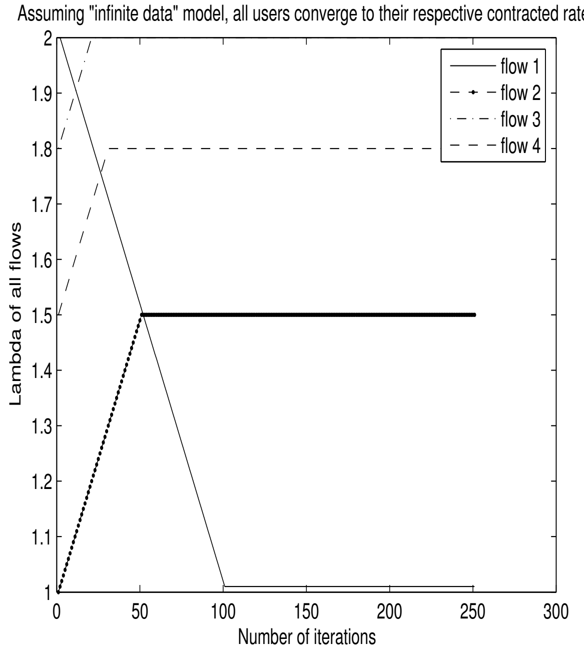

Figure 5: Schematic showing the convergence of data rates of all flows to their respective contracted rates.

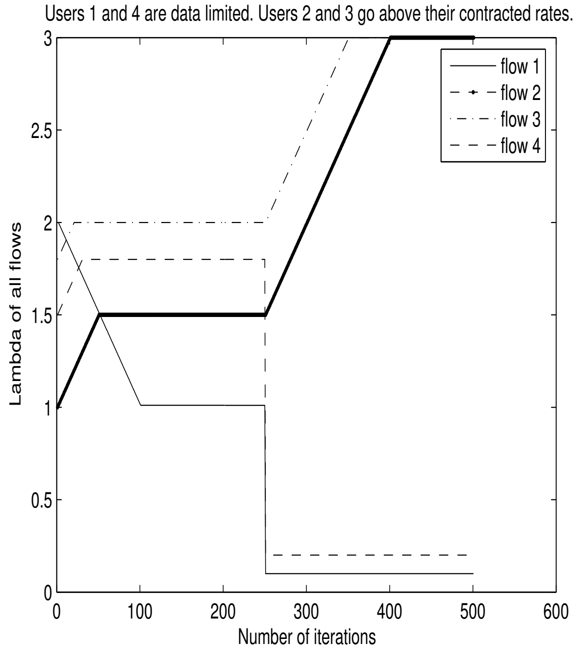

Figure 6: Schematic showing how users with enough data are encouraged to become overusers to utilize the spare network capacity.

Table 3: Routes of the flows after 100 iterations

	Link 1	Link 2	Link 3	Link 4	Link 5
User 1	1.00	1.00	0.00	0.00	0.00
User 2	0.00	0.00	1.49	0.00	0.00
User 3	1.99	0.00	0.00	0.00	1.99
User 4	0.00	0.00	1.79	0.00	1.79

Table 4: Routes of the flows after 400 iterations

	Link 1	Link 2	Link 3	Link 4	Link 5
User 1	0.0	0.0	0.1	0.1	0.0
User 2	0.0	0.0	3.0	0.0	0.0
User 3	3.0	0.0	0.0	0.0	3.0
User 4	0.0	0.0	0.2	0.0	0.2

Next, suppose that after 250 iterations, user 1 is data limited to a flow of 0.1 units and user 4 is data limited to 0.2 units. Now we want users 2 and 3 to go above their contracted rates and use the spare capacity left by underusers. From the routes used till now, we can say that user 2 can potentially increase up to $C_3 - \lambda_4 = 3.1$ and user 3 can increase up to $\min(C_5 - \lambda_4, C_1 - \lambda_1) = \min(3.6, 2.9) = 2.9$ units. In the simulation, we cap the flow rates of users 1 and 4 to 0.1 and 0.2 respectively and let users 2 and 3 vary rates to increase their net utilities as described earlier. Figure 6 shows that by the 400^{th} iteration, both users 2 and 3 increase flow to 3 units.

When we investigate the routes after 400 iterations (Table 4), we see that SP re-routes flow 1 through links 3 and 4. This allows user 2 to go up to $C_3 - \lambda_1 - \lambda_4 = 3$ units and user 3 to go up to $C_1 = 3$ units.

At the 500^{th} iteration, suppose that users 1 and 4 are back in the unlimited data phase. From this time onwards, we again let all users vary their rates. The pricing scheme compels overusers to back down, so that users 1 and 4 regain their shares. Figure 7 shows the simulation results. By the end of the 650^{th} iteration, all users converge back to their individual contracted rates.

This experiment illustrates that it *is* possible to assign prices once and for all at the beginning, such that this price vector ensures fair and efficient network sharing no matter what the sets of underusers and overusers may be. Also, the SP need not store I different prices for the I users.

5. STABILITY

When no user is constrained by limited amounts of available data, what is the rate vector that the collection of users converges to? If the increase in rate of any user causes a non-zero drop in the network, *i.e.*, due to increase in λ_i for any i, there is a $k = \{1, 2, \ldots, N\}$ and a $j = \{1, 2, \ldots, I\}$, such that $d^{(j)}(k) > 0$, we say that the network has no spare capacity. When prices are set according to Equation 8, $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_I^*)^t$ is a Nash equilibrium if, $\lambda_i^* \geq \gamma_i$ for all $i = \{1, 2, \ldots, I\}$ and there is no spare capacity in the network. This can be seen as follows. Considering any $i \in \{1, 2, \ldots, I\}$, we find that the user is either at contracted rate or already an overuser and the network is fully utilized. Therefore, Lemma 3.1 or 3.2 applies. This means that user i has no incentive to change her rate from λ_i^*, and this conclusion applies to all the users.

Hence, when prices are set in accordance with the conditions, system stability is assured. Now to calculate this λ^*,

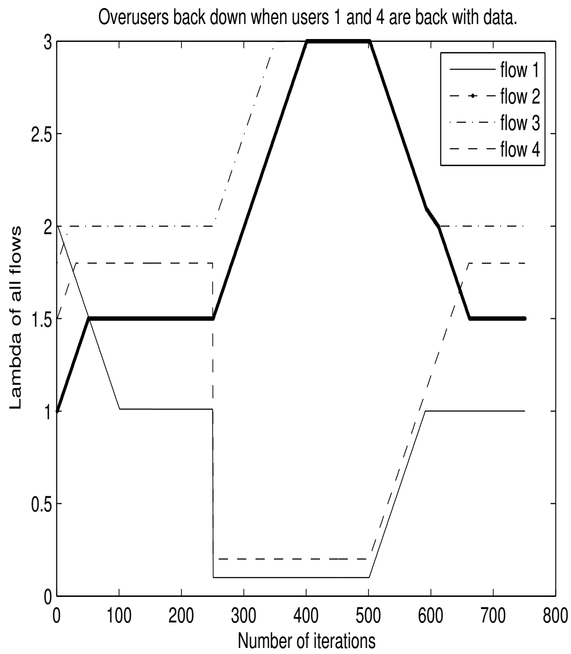

Figure 7: Schematic showing how the overusers come down to their contracted rates once underusers are back with data.

the set of equations which characterizes this system of users has to be solved. This system is as follows:

$$\frac{\mathrm{d}\lambda_i}{\mathrm{d}t} = \epsilon_i \mathrm{sgn}\Big(\frac{\partial NU_i(\vec{\lambda})}{\partial \lambda_i}\Big), \quad \forall\, i = \{1, 2, \ldots, I\} \tag{10}$$

where ϵ_i denotes the step size of user i and $sgn(x)$ is the function which is $+1$ if $x > 0$, -1 if $x < 0$ and 0 otherwise. After solving the maximization problem (9), if the SP finds that $d^{(i)}(k) > 0$ for some $k = \{1, 2, \ldots, N\}$, he infers that i is above her congestion threshold ct_i.

We have

$$\begin{aligned} NU_i(\vec{\lambda}) &= U_i\big(\lambda_i I_{\{\lambda_i \le ct_i\}} + \mu_i[\lambda_i] I_{\{\lambda_i > ct_i\}}\big) \\ &- D_i(\gamma_i - \lambda_i) I_{\{\lambda_i < \gamma_i\}} - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i I_{\{\lambda_i \le \gamma_i\}} \\ &- \pi_i P_i(\lambda_i - \gamma_i) I_{\{\lambda_i > \gamma_i\}} I_{\{\lambda_i > ct_i\}} \end{aligned} \tag{11}$$

where $I_{\{a \ge b\}}$ is the indicator function which takes a value of $+1$ if $a \ge b$ and 0 otherwise. $I_{\{\lambda_i > ct_i\}}$ can be evaluated depending upon the feedback from the SP. λ^* is an equilibrium point iff

$$\frac{\partial}{\partial \lambda_i} NU_i(\vec{\lambda})\bigg|_{\lambda=\lambda^*} = 0, \quad \forall\, i = \{1, 2, \ldots, I\} \tag{12}$$

Given an initial flow vector and the step sizes ϵ_i's, we numerically solve this system of coupled partial differential equations and find out the equilibrium point.

6. CONCLUSION

We considered sources that could occasionally be constrained by limited amounts of available data. Further, each user has a contract with the SP, specifying the rate at which she can send traffic into the network. We were interested in a pricing scheme that would ensure fair and efficient sharing of the network resources.

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 operation is possible and observed that, under conditions that can be easily ensured, *one* 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 network. A simple experiment in MATLAB demonstrated the utility of our approach.

We recognize the following limitation of our work. We have not explicitly considered the problem of revenue maximization for the SP. While our goals of fair and efficient sharing of the network are natural, we would like to consider the problem of explicit revenue maximization as well.

7. REFERENCES

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

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

[3] F. Kelly. Charging and rate control for elastic traffic (corrected version). *European Transaction on Telecommunication*, 8(1):33–37, Jan 1997.

[4] F. Kelly, A. K. Maulloo, and D. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *J. Oper. Res. Soc.*, 49(3):237–252, Mar 1998.

[5] S. Kunniyar and R. Srikant. End-to-end congestion control schemes: Utility functions, random losses and ecn marks. *IEEE/ACM Transactions on Networking*, 11(5):689–702, Oct 2003.

[6] R. J. La and V. Anantharam. Utility-based rate control in the internet for elastic traffic. *IEEE Transactions on Networking*, 10(2):272–286, Apr 2002.

[7] 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.

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

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

[10] P. Marbach. Priority service and max-min fairness. *IEEE/ACM Transactions on Networking*, 11(5):733–746, Oct 2003.

[11] P. Marbach. Analysis of a static pricing scheme for priority services. *IEEE/ACM Transactions on Networking*, 12(2):312–325, Apr 2004.

[12] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, Oct 2000.

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

APPENDIX

Proof of Lemma 3.1

Since congestion threshold is higher than the contracted rate, there cannot be any congestion in the network as long as $\lambda_i < \gamma_i$. So, the total flow injected by user i is completely carried by the network thereby making $\mu_i = \lambda_i$. The net utility function for user i is:

$$NU_i = U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i$$

User i has incentive to increase her rate if and only if her net utility increases with increase in λ_i. That is,

$$\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i} \geq 0$$

$$U_i'(\lambda_i) + D_i'(\gamma_i - \lambda_i) - \Big(\frac{\pi_i}{\gamma_i}\Big) \geq 0$$

For NU_i to keep increasing till user i reaches γ_i, it is sufficient to have

$$\pi_i \quad \leq \quad \gamma_i(U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x))$$

This condition is also necessary because if $\pi_i > \gamma_i(U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x))$, then there exists a $\lambda_i < \gamma_i$ such that $\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i} < 0$. This can be seen as follows. If $\pi_i = \gamma_i(U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x)) + \epsilon$, where $\epsilon > 0$, then $\frac{\pi_i}{\gamma_i} = U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x) + \frac{\epsilon}{\gamma_i}$. Now,

$$\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i}\bigg|_{\lambda_i\uparrow\gamma_i} \quad = \quad -\frac{\epsilon}{\gamma_i}$$

When λ_i is sufficiently close to γ_i, $\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i}$ becomes negative. Thus, user i looses incentive to increase rate even before reaching her contracted rate.

Since congestion threshold is above contracted rate, user i should have incentive to increase her rate further till the network is fully utilized, *i.e.*, till λ_i is equal to ct_i. When $\lambda_i \in [\gamma_i, ct_i]$, NU_i is simply $U_i(\lambda_i) - \pi_i$ which is an increasing function of λ_i. So, user i will have no problem in increasing λ_i further.

Once i shoots above ct_i, the network gets congested and there is extra penalty. Therefore,

$$NU_i \quad = \quad U_i(\mu_i[\lambda_i]) - \pi_i P_i(\lambda_i - \gamma_i)$$

$\mu_i[\lambda_i]$ can be obtained numerically by solving the linear system of flow conservation and capacity constraint equations, given the injected traffic of all users. One observation to make here is that $\mu_i[ct_i] = ct_i$ because, at $\lambda_i = ct_i$ there is no congestion. We can prove that as long as we choose $\pi_i \geq 0$ and $P_i(ct_i - \gamma_i) \geq 1$, the net utility for user i at a rate just above ct_i is less than the net utility at ct_i. Since $P_i(x)$ is an convex increasing function of x, and $P_i(0) = 1$, $P_i(ct_i - \gamma_i) > 1$. Apart from this, we want to ensure that NU_i decreases with λ_i when λ_i is greater than ct_i. For this to happen,

$$\frac{\mathrm{d}NU_i}{\mathrm{d}\lambda_i} \leq 0$$

when $\lambda_i > ct_i$. Let $\lambda_i = ct_i + \delta$, for some $\delta > 0$. Then, by definition

$$NU_i \quad = \quad U_i(\mu_i[ct_i + \delta]) - \pi_i P_i(ct_i + \delta - \gamma_i)$$

and we want $\frac{\mathrm{d}NU_i}{\mathrm{d}\delta} \leq 0$. On simplification we get,

$$\pi_i \geq \frac{U_i'(\mu_i[ct_i + \delta])\mu_i'[ct_i + \delta]}{P_i'(ct_i + \delta - \gamma_i)}$$

Because we want the lower bound to hold for *every* $\delta > 0$, we now take the supremum of the lower bound over all $\delta > 0$. Let us call value of δ which maximizes the RHS as δ^* This yields

$$\pi_i \geq \frac{U_i'(\mu_i[ct_i + \delta^*])\mu_i'[ct_i + \delta^*]}{P_i'(ct_i + \delta^* - \gamma_i)}$$

We have proved that the lower bound on π_i is sufficient. Necessity is shown as follows. Let $\epsilon > 0$ be a given small number. If we take $\big(\frac{U_i'(\mu_i[ct_i+\delta^*])\mu_i'[ct_i+\delta^*]}{P_i'(ct_i+\delta^*-\gamma_i)} - \epsilon\big)$ 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. □

Proof of Lemma 3.2

There is no congestion in the network and no packets are dropped unless user i overshoots her contracted rate. The expression for net utility is,

$$NU_i = U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i$$

Following similar algebra as done in the proof of Lemma 3.1, we can show that the necessary and sufficient condition for user i to increase her rate up to γ_i is

$$\pi_i \leq \gamma_i(U_i'(\gamma_i) + \lim_{x\to 0+} D_i'(x))$$

Since congestion threshold is equal to the contracted rate, once user i reaches her contracted rate, the network is full. So, she should have no more incentive to increase rate any further. From the definition of net utility function it is clear that when $\pi_i \geq 0$ and $P_i(0) \geq 1$, NU_i at $\lambda_i = \gamma_i$ is greater than NU_i at $\lambda_i > \gamma_i$. Since NU_i should be decreasing with λ_i once $\lambda_i > \gamma_i$,

$$\frac{\mathrm{d}}{\mathrm{d}\delta}\Big(U_i(\mu_i[\gamma_i + \delta]) - \pi_i P_i(\delta)\Big) \quad \leq \quad 0$$

This can be reduced to

$$\pi_i \quad \geq \quad \max_{\delta\geq 0}\Big(\frac{U_i'(\mu_i[\gamma_i + \delta])\mu_i'[\gamma_i + \delta]}{P_i'(\delta)}\Big)$$

Let δ^* be the value of δ that maximizes the RHS. Then

$$\pi_i \quad \geq \quad \frac{U_i'(\mu_i[\gamma_i + \delta^*])\mu_i'[\gamma_i + \delta^*]}{P_i'(\delta^*)}$$

This proves that the lower bound on π_i is sufficient. The necessity of the lower bound can be seen in exactly the same way as done in the case of previous Lemma. □

Proof of Lemma 3.3

Since congestion threshold is below contracted rate, user i has room to increase her rate only up to ct_i without causing congestion in the network. This is possible if,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_i}\Big(U_i(\lambda_i) - D_i(\gamma_i - \lambda_i) - \Big(\frac{\lambda_i}{\gamma_i}\Big)\pi_i\Big) \quad \geq \quad 0$$

This NU_i should not cease to increase till λ_i reaches ct_i. We can simplify the above to get following sufficient condition on π_i.

$$\pi_i \leq \gamma_i\big(U_i'(ct_i) + D_i'(\gamma_i - ct_i)\big) \qquad (13)$$

Since $ct_i < \gamma_i$, we want i to increase her rate further up to her rightful share of γ_i. So, for $0 \le \delta \le (ct_i - \gamma_i) = \eta_i$,

$$\frac{\mathrm{d}}{\mathrm{d}\delta}\left(U_i(\mu_i[ct_i + \delta]) - D_i(\eta_i - \delta) - \left(\frac{ct_i + \delta}{\gamma_i}\right)\pi_i\right) \ge 0$$

which reduces to

$$\pi_i \le \min_{0 \le \delta \le \eta_i} \gamma_i\left(U_i'(\mu_i[ct_i + \delta])\mu_i'[ct_i + \delta] + D_i'(\eta_i - \delta)\right)$$

Let RHS be minimized at $\delta = \delta^+$. Then,

$$\pi_i \le \gamma_i\left(U_i'(\mu_i[ct_i+\delta^+])\mu_i'[ct_i+\delta^+]+D_i'(\gamma_i-ct_i-\delta^+)\right) \quad (14)$$

To prove that both the upper bounds (Equation 13 and 14) are also necessary, we can proceed on similar lines as done in Lemma 3.1.

Now, NU_i at $\lambda_i = \gamma_i$ should exceed NU_i at $\lambda_i > \gamma_i$ and NU_i should keep reducing with λ_i when $\lambda_i \in (\gamma_i, \infty)$. Mathematically, for a small $\delta > 0$,

$$U_i(\gamma_i) - \pi_i \quad \ge \quad U_i(\mu_i[\gamma_i + \delta]) - \pi_i P_i(\delta)$$

which can be ensured if $\pi_i \ge 0$ and $P_i(0) \ge 1$. Also,

$$\frac{\mathrm{d}}{\mathrm{d}\delta}\left(U_i(\mu_i[\gamma_i + \delta]) - \pi_i P_i(\delta)\right) \le 0$$

which implies

$$\pi_i \quad \ge \quad \frac{U_i'(\mu_i[\gamma_i + \delta^*])\mu_i'[\gamma_i + \delta^*]}{P_i'(\delta^*)}$$

where $\delta^* = \arg\max_{\delta \ge 0}\left(\frac{U_i'(\mu_i[\gamma_i+\delta])\mu_i'[\gamma_i+\delta]}{P_i'(\delta)}\right)$. The necessity of this bound can be proved by contradiction following similar steps as done in previous cases. □