A Pricing Model for High Speed Networks with Guaranteed Quality of Service*

Hong Jiang and Scott Jordan
hcs@eecs.nwu.edu & scott@eecs.nwu.edu

Department of Electrical Engineering & Computer Science
Northwestern University
2145 Sheridan Rd.
Evanston, IL 60208-3118

Abstract

In this paper, we discuss the role of prices in combining user characterization, network resource allocation, and contract negotiation to form a complete connection establishment process. We suggest that such a process should encourage network efficiency through distributed resource allocation among virtual circuits, circuit bundles, and virtual paths. We adopt effective bandwidth as our user traffic characterization and our pricing base, and we measure network efficiency by total user benefit. We allow a limited degree of statistical multiplexing by incorporating multiplexing gain into the prices. Finally, we propose a hierarchical and distributed negotiation structure under which only hierarchically adjacent and geographically local network entities communicate with each other.

1 Introduction

In this paper, we set up a mathematical framework to investigate the role of prices in the connection establishment process. We suggest that prices can allocate resources at each of three network levels in a manner that encourages network efficiency, as measured by total user benefit.

Some recent research has investigated the use of pricing in implementing a distributed contract negotiation process. Best-effort service computer networks have been studied in [1] [2] [3], and ATM networks have been considered in [4] [5] [6]. Although pricing methods for best-effort service can not be directly used in networks which guarantee QoS, the ideas are similar. The ATM work typically focuses on specific network processes. For example, Kelly [4] devised a pricing structure to encourage users to reveal the true mean rates of their traffic streams so that mean-rate policing is unnecessary. Low [5] exploited the trade-off between buffer and bandwidth to improve network efficiency, and Murphy [6] emphasized issues of resources allocation among different network levels.

In [7], we reviewed recent contributions to each of these steps, and suggested problems that must be solved to integrate these into a complete connection establishment process. Our pricing scheme in this paper is based on the framework laid in [7], where effective bandwidth is used as both the user traffic characterization and the pricing base. Our method differs from others in that certain degree of statistical multiplexing is allowed and the effects of multiplexing gain on resource allocation and prices are studied. Furthermore, we propose a hierarchical and distributed negotiation structure under which only adjacent and local network levels communicate with each other.

In Section 2, the framework of a connection establishment process is described. User traffic streams are characterized by effective bandwidth and user's valuation of service is defined by benefit functions. The network's capabilities to offer a particular service mix according to the partition of resources at three levels is also described. The structure of the negotiation processes are explained. In Section 3, mathematical models are established based on the framework described in previous sections. Section 4 studies the interaction among multiplexing gains, prices and costs for identical users. Finally, in the last section, we discuss remaining issues concerning design and convergence of iterative negotiation processes.

2 Connection Establishment Process

2.1 User Characterization

We assume a user's objective is to maximize her consumer surplus, defined as, the difference between benefit and cost.

* This work was supported by the Ameritech foundation.
We model a user's demand for bandwidth for a real-time service through user benefit as a function of pre-transmission loss\(^1\). For example, a user might adjust the compression levels before her video transmission according to prices for bandwidth.

We model a user's demand for bandwidth for a non-real-time service through user benefit as a function of completion time. For example, a user can decide to transmit her data at once or to spread out the data transmission, depending on prices for bandwidth. Benefit functions for non-real-time service could change with elapsed time and remaining file size.

We choose effective bandwidth to characterize user traffic streams. Effective bandwidth (e.g., [8]) models a source's use of network resources at the user/network interface. The effective bandwidth is bounded by the source's mean and peak rates, and is a function of the source's burstiness and of 
\[
\zeta = \frac{\log (\text{loss probability})}{\text{buffer length}}.
\]
Effective bandwidths of the traffic sources sharing the same buffer are additive and are independent of other sources.

For real-time service, effective bandwidth is calculated for the transmitted traffic stream given a channel's QoS defined by cell loss probability and maximal delay jitter. Thus we obtain effective bandwidth as a function of pre-transmission loss to characterize user traffic streams.

For non-real-time service, if a user transmits the data at a constant rate during each negotiation cycle, the effective bandwidth of the traffic is equal to its mean rate. Effective bandwidth is inversely proportional to the completion time.

2.2 Network Characterization

We consider the case where the network is a public good and its objective is to maximize total user benefit of all network users [1] [5]. As a result of our network objective, the price is set to zero when a channel is under-utilized and thus the revenue collected does not guarantee cost recovery or profitability. Many papers, however, have addressed these telecommunications networks issue using schemes such as peak-load pricing and two-part tariffs and found that dynamic pricing plus a flat fee can often both achieve network efficiency and recover cost (c.f. [9]).

We explicitly characterize a network's current capabilities to offer a particular service mix according to the partition of resources at three levels: virtual circuit (VC), circuit bundle (CB), and virtual path (VP). Among many types of resource allocations [7], VP/QoS allocation is adopted in this paper, where all VCs with identical paths and QoS are statistically multiplexed.

Effective bandwidth is used as a tool for admission control. Under a VP/QoS architecture, we choose a particularly simple admission control policy: a new call request can be accepted if the sum of effective bandwidths of all users (including the new caller) within its chosen circuit bundle is no more than the circuit bundle's capacity.

2.3 Contract Negotiation

The key to efficient network usage is to charge users fairly according to the amount of resources occupied by individual users. Effective bandwidth of a traffic stream can be considered as composed of two parts: mean rate + burstiness. We therefore propose that users be charged an amount equal to price \( \times \) effective bandwidth. A traffic stream will thus be charged for its mean rate plus an amount based on its burstiness.

Contract negotiation is performed at three levels and at different time scales with the lowest level having the shortest negotiation cycle. Each network level maximizes its total user benefit as a supplier as it negotiates with adjacent lower network levels by setting the right prices; and maximizes its consumer surplus as a consumer as it negotiates with adjacent higher level by choosing the right demand. For example, when negotiation takes place between a virtual path and its circuit bundles, the virtual path is a supplier while its circuit bundles are consumers. For a negotiation between a circuit bundle and its virtual circuits, the role of the circuit bundle is switched to a supplier while its virtual circuits become the consumers. Consequently, computations of optimal resource allocations are distributed among network levels and among local network areas.

3 Mathematical Models

In this section, we explicitly set up the mathematical models for optimal resource allocation and pricing based on the framework developed in the last section. We assume the routing of virtual paths is fixed and there is only one path from a source to a destination. Terms "user \((i, j, k)\)" and "virtual circuit \((i, j, k)\)" are used interchangeably.

Under the VP/QoS allocation scheme, virtual circuits sharing the same virtual path and QoS are statistically multiplexed within a circuit bundle, and the statistical multiplexing gain is considered in the admission control and pricing. Since we assume that non-real-time data can be transmitted at a constant rate during a negotiation cycle, no statistical multiplexing gain exists for non-real-time traffic. Due to the variable-rate traffic of real-time service, statistical multiplexing gain is present.

Notation is specified as follows:

- \(E_{i,j,k}\): effective bandwidth of virtual circuit (VC) \(i\) within circuit bundle (CB) \(j\) which is within virtual path (VP) \(k\).
- \(S_{j,k}\): capacity of CB \(j\) within VP \(k\).
- \(T_i\): capacity of trunk \(i\).
- \(V_k\): capacity of VP \(k\).
- \(Q_{i,j,k}\): pre-transmission loss for real-time services and completion time for non-real-time services.
- \(b_{i,j,k}\): benefit function for VC \(i\) within CB \(j\) and within VP \(k\), or benefit function of user \((i,j,k)\).

---

\(^1\) Pre-transmission loss refers to the loss incurred before transmission.
\[
ben_{jk} = \sum_i^k \text{aggregate benefit function of circuit bundle } (j,k).
\]

\[
ben_k = \sum_j^k \sum_i^k \text{aggregate benefit function of virtual path } k.
\]

\[
I_{kl} = \begin{cases} 
1 & \text{if VP } k \text{'s route uses trunk } l, \text{ } I_{kl} \text{ is equal to } 1. \text{ Otherwise, it is equal to } 0.
\end{cases}
\]

\[
\lambda_{jk}; \text{ the Lagrangian multiplier and the unit price in $/bit/second of } CB \ j \text{ within VP } k.
\]

\[
\mu_k; \text{ the Lagrangian multiplier and the unit price in $/bit/second of VP } k.
\]

\[
\rho_l; \text{ the Lagrangian multiplier and the unit price in $/bit/second of trunk } l.
\]

Cell loss probability and delay jitter of a circuit bundle are assumed to be fixed. We assume that the effective bandwidth of a real-time service is decreasing, differentiable and jointly convex in a circuit bundle’s capacity $S_{ij}$ and pre-transmission loss $Q_{ijk}$ for a fixed QoS. Higher pre-transmission loss results in less effective bandwidth. Higher capacity allows a longer buffer for the same maximal delay jitter, and therefore it reduces the effective bandwidth needed for the same cell loss probability. For non-real-time services, we assume that effective bandwidth is decreasing, differentiable and convex in completion time $Q_{ijk}$, but does not change with capacity $S_{ij}$. For convenience, we denote the effective bandwidth for both types of service by $E_{ijk}(Q_{ijk}, S_{jk})$. In addition, we assume that benefit functions are concave, decreasing and differentiable in $Q_{ijk}$.

### 3.1 A Centralized Model

Assuming that the network knows its trunk capacities and virtual path routing, and every user’s benefit function and traffic stream characterization, the network can perform total user benefit maximization for fixed routing and trunk sizes:

\[
\begin{align*}
\text{max} & \quad Q_{ijk}, S_{jk}, V_k \sum_i^j \sum_k^j \text{ben}_{ijk}(Q_{ijk}) \\
\text{subject to constraints:} & \\
\sum_j^k E_{ijk}(Q_{ijk}, S_{jk}) & \leq S_{jk} \quad \forall (j, k) \\
\sum_k^j S_{jk} & \leq V_k \quad \forall k \\
\sum_k^j V_k T_l & \leq T_l \quad \forall l \\
Q_{ijk}, S_{jk}, V_k & \geq 0 \quad \forall (i, j, k)
\end{align*}
\]

The Lagrangian function is:

\[
L(\lambda_{jk}, \mu_k, \rho_l) = \sum_i^j \sum_k^j \text{ben}_{ijk}(Q_{ijk}) + \sum_j^k \lambda_{jk} \cdot \\
\left( S_{jk} - \sum_i^j E_{ijk} \right) + \sum_k^j \mu_k \cdot \left( V_k - \sum_j^k S_{jk} \right) + \sum_l^j \rho_l \cdot \left( T_l - \sum_k^j V_k T_l \right)
\]

Constraint function $S_{jk} - \sum_i^j E_{ijk}(Q_{ijk}, S_{jk})$ is a concave function since $\sum_i^j E_{ijk}(Q_{ijk}, S_{jk})$ is convex.

Because the objective function and the constraint functions in (6) are concave, this is a concave program and Kuhn-Tucker conditions [10] are sufficient and necessary for the global optimal solution. Kuhn-Tucker theorem gives (2)-(4) plus the following conditions for optimality:

\[
\frac{\partial}{\partial Q_{ijk}} L(\lambda_{jk}, \mu_k, \rho_l) = \frac{\partial}{\partial Q_{ijk}} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk} \frac{\partial E_{ijk}}{\partial Q_{ijk}} \leq 0 \quad \forall (i, j, k)
\]

(7)

\[
\frac{\partial}{\partial S_{jk}} L(\lambda_{jk}, \mu_k, \rho_l) = \frac{\partial}{\partial S_{jk}} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk} \frac{\partial E_{ijk}}{\partial S_{jk}} = 0 \quad \forall (j, k)
\]

(8)

\[
\frac{\partial}{\partial V_k} L(\lambda_{jk}, \mu_k, \rho_l) = \frac{\partial}{\partial V_k} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk} \sum_l^j V_k T_l - \mu_k \leq 0 \quad \forall l
\]

(9)

(10)

(11)

(12)

(13)

(14)

(15)

Solving such a maximization problem for a network of moderate size can be computationally intensive, and thus it is desirable to distribute the computation to various network levels and local network areas. If we consider the Lagrangian multipliers $\lambda_{jk}, \mu_k, \rho_l$ to be the prices charged by CB $(j, k)$, VP $k$, and trunk $l$ respectively, equations (7)-(8), (9)-(10), and (11)-(12) describe VC $(i, j, k)$, CB $(j, k)$, and VP $k$’s behavior in response to the prices as a consumer, and equations (2)-(4), (13)-(15) describe the VC, CB and VP’s strategy as a supplier. By decomposing the Kuhn-Tucker conditions into separate roles of consumer and supplier at each network level, we might be able to transform the centralized problem into a distributed problem.

Note that revenue is only collected by the CB level since the CB price $\lambda_{jk}$ is the real price charged to its users. The VP price $\mu_k$ and trunk price $\rho_l$ are the internal prices to signal the optimality of resource allocation among the CB, VP and trunk levels. Equations (13)-(15) are the complementary slackness conditions which indicate if the second terms in these equations are strictly less than zero, then the first terms (Lagrangian multipliers) have to be
zero. Put in the context of network resource allocation and pricing, these equations mean that if the channel capacity at a network level is not fully utilized, then the channel should set its price to zero as a supplier. This is a characteristic of public goods which are allocated to maximize total user benefit. As mentioned above, the pricing structure can be modified to guarantee cost recovery.

3.2 A Distributed Model

Network maximization is distributed to and performed at three network levels: user - circuit bundle, circuit bundle - virtual path, virtual path - physical trunk. Each network level adjusts the price to make the demand and supply meet, otherwise it sets the price to zero. When an equilibrium price is reached, the total user benefit function or the sensitivity of it is passed on to the adjacent higher network level, and the capacity allocated to the network level is passed on to the adjacent lower network level.

- User - Circuit Bundle Negotiation

User \((i, j, k)\) maximizes his consumer surplus given his circuit bundle's price \(\lambda_{jk}\) and the capacity of the circuit bundle \(S_{jk}\):

\[
\max_{Q_{ijk}} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk}E_{ijk}(Q_{ijk}, S_{jk}), \quad \text{subject to } Q_{ijk} \geq 0
\]

(17)

The optimal solution satisfies:

\[
\frac{\partial}{\partial Q_{ijk}} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk} = 0
\]

(18)

\[
Q_{ijk} \left( \frac{\partial}{\partial Q_{ijk}} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk} \right) = 0
\]

(19)

(18) and (19) show that if the marginal benefit \((-\lambda_{jk})\) is strictly greater than the marginal cost of effective bandwidth \(-E_{ijk}\) for all values of pre-transmission loss or completion time \(Q_{ijk}\), then set \(Q_{ijk} = 0\). If the two can be equal, then the optimal \(Q_{ijk}\) can be solved from:

\[
\frac{\partial}{\partial Q_{ijk}} \text{ben}_{ijk}(Q_{ijk}) - \lambda_{jk} = 0
\]

(20)

(20) means that the marginal benefit equals the marginal cost at \(Q_{ijk}\) for all users who choose to use the service.

Circuit bundle \((j, k)\) maximizes its total user benefit given fixed capacity \(S_{jk}\) as a supplier:

\[
\max_{Q_{ijk}} \sum_i \text{ben}_{ijk}(Q_{ijk})
\]

subject to constraints:

\[
\sum_i E_{ijk}(Q_{ijk}, S_{jk}) \leq S_{jk}
\]

(22)

\[
Q_{ijk} \geq 0
\]

(23)

The Lagrangian function is:

\[
L_{jk}(Q_{ijk}, \lambda_{jk}) = \sum_i \text{ben}_{ijk}(Q_{ijk}) + \lambda_{jk} \left( S_{jk} - \sum_i E_{ijk} \right)
\]

where \(\lambda_{jk} \geq 0\)

(24)

Kuhn-Tucker theory indicates that if \(Q'_{ijk}, \lambda'_{jk}\) solve the saddle-value problem of (25), then \(Q'_{ijk}\) solves the above maximization problem.

\[
L_{jk}(Q_{ijk}, \lambda_{jk}) = \max_{Q_{ijk}} \min_{\lambda_{jk}} L_{jk}(Q_{ijk}, \lambda_{jk})
\]

(25)

Equation (25) says that the tasks of optimization can be performed separately by the circuit bundle and its users. Circuit bundle \((j, k)\) minimizes total user benefit of the circuit bundle \(L_{jk}\), for fixed demands \(Q_{ijk}\), by varying the circuit bundle's price \(\lambda_{jk}\), which leads to

\[
S_{jk} - \sum_i E_{ijk}(Q_{ijk}, S_{jk}) \geq 0
\]

(26)

\[
\lambda_{jk} \left( S_{jk} - \sum_i E_{ijk}(Q_{ijk}, S_{jk}) \right) = 0
\]

(27)

On the other hand, user \((i, j, k)\) maximizes \(L_{jk}\), for a fixed circuit bundle price \(\lambda_{jk}\), by varying its demand \(Q_{ijk}\), which results in equations (18)-(19). These equations mean that users will behave in a socially optimal way as they maximize their own consumer surplus if the right prices are charged, assuming that collusion and market manipulation by users are absent. Circuit bundles do not need to collect each user's benefit function if the right prices can be obtained through some iterative negotiation.

Intuitively, a user's pre-transmission loss \(Q_{ijk}\) decreases as his circuit bundle's capacity \(S_{jk}\) increases since larger channel capacity allows more multiplexing gain and reduces the price so that the user could afford to transmit more data.

**Theorem 1:**

A user's pre-transmission loss for real-time services or completion time for non-real-time services \(Q_{ijk}\) is convex and decreasing in his circuit bundle's capacity \(S_{jk}\).

**Proof:**

To prove that \(Q_{ijk}\) is decreasing in \(S_{jk}\), we start by taking the derivatives with respect to \(S_{jk}\) on both sides of (28) and obtain (29).

\[
\sum_i E_{ijk}(Q_{ijk}, S_{jk}) = S_{jk} \quad \forall (j,k)
\]

(28)

\[
\sum_i \left( \frac{\partial E_{ijk}}{\partial Q_{ijk}} \frac{\partial Q_{ijk}}{\partial S_{jk}} + \frac{\partial E_{ijk}}{\partial S_{jk}} \right) = 1 \quad \forall (j,k)
\]

(29)
From our assumption that effective bandwidth $E_{ijk}$ is decreasing in both $Q_{ijk}$ and $S_{jk}$, we know that $\frac{\partial E_{ijk}}{\partial Q_{ijk}}$ and $\frac{\partial E_{ijk}}{\partial S_{jk}}$ are negative for all users. Since we have also assumed that benefit functions $ben_{ijk}$ are concave and increasing for all users, $\frac{\partial Q_{ijk}}{\partial S_{jk}}$ has the same sign for all users. Thus, we know that $\frac{\partial E_{ijk}}{\partial Q_{ijk}} \frac{\partial Q_{ijk}}{\partial S_{jk}} + \frac{\partial E_{ijk}}{\partial S_{jk}}$ has the same sign for all users, then (29) implies (30).

$$\frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} = \frac{\partial^2 E_{ijk}}{\partial S_{jk}^2} \geq 0 \quad \forall (i, j, k)$$

(30)

Since $\frac{\partial E_{ijk}}{\partial S_{jk}} \leq 0$ and $\frac{\partial Q_{ijk}}{\partial S_{jk}} \leq 0$ from our assumption, we know that $\frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} \leq 0$ is true.

When the CB's capacity is not fully utilized, then $\frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} = 0$. To prove the convexity, we need to show that $\frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} \geq 0$ when the capacity is fully utilized. Assuming $\frac{\partial^2 E_{ijk}}{\partial S_{jk}^2} = \frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2}$, we obtain (31) by taking the derivatives with respect to $S_{jk}$ on both sides of (29).

$$\sum_i \frac{\partial^2 E_{ijk}}{\partial Q_{ijk}^2} \frac{\partial Q_{ijk}^2}{\partial S_{jk}^2} + \frac{\partial E_{ijk}}{\partial Q_{ijk}^2} \frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} + \frac{\partial^2 E_{ijk}}{\partial S_{jk}^2} = 0$$

(31)

Since $E_{ijk}$ is jointly convex in $Q_{ijk}$ and $S_{jk}$, (32) is true for any real numbers $A$ and $B$.

$$\frac{\partial^2 E_{ijk}}{\partial Q_{ijk}^2} A^2 + 2AB \frac{\partial^2 E_{ijk}}{\partial Q_{ijk} \partial S_{jk}} + B^2 \frac{\partial^2 E_{ijk}}{\partial S_{jk}^2} \geq 0 \quad \forall (i, j, k)$$

(32)

Combine (32) with (31), we get

$$\sum_i \frac{\partial E_{ijk}}{\partial Q_{ijk}} \frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} \leq 0 \quad \forall (i, j, k)$$

(33)

Again from previous assumptions about users' behavior, $\frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2}$ has the same sign for all users, and (33) and $\frac{\partial E_{ijk}}{\partial Q_{ijk}} \leq 0$ give (34).

$$\frac{\partial^2 Q_{ijk}}{\partial S_{jk}^2} \geq 0 \quad \forall (i, j, k)$$

Q.E.D

When the capacity is fully utilized, we can solve for the optimal solutions $Q_{ijk}$ and $\lambda'_{jk}$ in terms of the circuit bundle's capacity $S_{jk}$. The envelope theorem indicates that:

$$\frac{d}{dS_{jk}} \text{ben}_{ijk}(S_{jk}) = \frac{\partial L_{jk}}{\partial S_{jk}} \bigg|_{\lambda_{jk} = \lambda'_{jk}(S_{jk})} - \lambda'_{jk}(S_{jk})$$

$$= \lambda'_{jk}(S_{jk}) \left( 1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk}(Q_{ijk}(S_{jk}), S_{jk}) \right)$$

(35)

If the circuit bundle's capacity is not fully utilized, then the optimal price per effective bandwidth unit $\lambda'_{jk} = 0$, and thus $\frac{d}{dS_{jk}} \text{ben}_{ijk}(S_{jk}) = 0$.

**Lemma 1:**

If $f(g)$ is concave and decreasing, and if $g(x)$ is convex, then $f(x)$ is concave in $x$.

Next, we would like to prove the concavity of the aggregate benefit function of a circuit bundle $\text{ben}_{jk}(S_{jk})$ using Theorem 1 and Lemma 1.

**Theorem 2:**

The aggregate benefit function of a circuit bundle $\text{ben}_{jk}(S_{jk})$ is concave and increasing in its capacity $S_{jk}$.

**Proof:**

$\text{ben}_{jk}(S_{jk})$ is increasing due to the fact that $\lambda'_{jk}$ and $1 - \sum_i \frac{\partial E_{ijk}}{\partial S_{jk}}$ are nonnegative in (35). When the CB's capacity is not fully utilized, $\frac{d}{dS_{jk}} \text{ben}_{jk} = 0$. We need to prove that $\frac{d^2}{dS_{jk}^2} \text{ben}_{jk} \leq 0$ when the capacity is fully utilized.

$\text{ben}_{jk}(Q_{ijk})$ is a composite function of $\text{ben}_{jk}(Q_{ijk})$ and $Q_{ijk}(S_{jk})$. We know that $\text{ben}_{jk}(Q_{ijk})$ is concave and decreasing in $Q_{ijk}$ from the previous assumption, and Theorem 1 shows that $Q_{ijk}$ is convex in $S_{jk}$. Therefore Lemma 1 leads to Theorem 2.

Q.E.D

- **Circuit bundle - Virtual path Negotiation**

From Theorem 2, we know that benefit function $\text{ben}_{jk}(S_{jk})$ is concave and increasing in circuit bundle's capacity $S_{jk}$. The maximization problem between a circuit
bundle and a virtual path is again a concave program. If we can obtain the sensitivity of the benefit function \( \text{ben}_{jk}(S_{jk}) \), e.g. by (35), then the circuit bundle is able to maximize its consumer surplus given its VP's price \( \mu_k \). Circuit bundle \((j,k)\) performs:

\[
\max_{S_{jk}} \text{ben}_{jk}(S_{jk}) - \mu_k S_{jk}, \quad \text{subject to } S_{jk} \geq 0
\]  

(36)

The optimal solution satisfies

\[
\frac{\partial}{\partial S_{jk}} \text{ben}_{jk}(S_{jk}) - \mu_k \leq 0
\]  

(37)

\[
S_{jk} \left( \frac{\partial}{\partial S_{jk}} \text{ben}_{jk}(S_{jk}) - \mu_k \right) = 0
\]  

(38)

Virtual path \( k \) maximizes its total user benefit given its fixed capacity \( V_k \):

\[
\max_{S_{jk}} \sum_j \text{ben}_{jk}(S_{jk})
\]  

subject to constraints:

\[
\sum_j S_{jk} \leq V_k
\]  

(40)

\[
S_{jk} \geq 0
\]  

(41)

The Lagrangian function is:

\[
L_k(S_{jk}, \mu_k) = \sum_j \text{ben}_{jk}(S_{jk}) + \mu_k \left( V_k - \sum_j S_{jk} \right)
\]  

where \( \mu_k \geq 0 \)

(42)

Using the same saddle-value argument, virtual path \( k \) obtains (43)-(44), and circuit bundle \((j,k)\) obtains (37)-(38), which means that the total user benefit of virtual path \( k \) is maximized as circuit bundles maximize their own consumer surplus and that virtual path \( k \) does not need to know its circuit bundles' benefit functions.

\[
\mu_k \left( V_k - \sum_j S_{jk} \right) = 0 \quad \forall k
\]  

(43)

\[
V_k - \sum_j S_{jk} \geq 0 \quad \forall k
\]  

(44)

Substituting the sensitivity of each bundle's benefit to allocated bandwidth given by (35) into (36) and (37), we get

\[
\lambda_{jk} \left( 1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk} Q_{ijk}(S_{jk}, S_{jk}) \right) - \mu_k \leq 0 \quad \forall (j,k)
\]  

(45)

\[
S_{jk} \lambda_{jk} \left( 1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk} Q_{ijk}(S_{jk}, S_{jk}) \right) - \mu_k = 0 \quad \forall (j,k)
\]  

(46)

This complementary slackness condition indicates that if the second term in (46) is less than zero, then no capacity will be allocated to CB \((j,k)\). If the second term is equal to zero, then (47) has to be true.

\[
\lambda_{jk} \left( 1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk} Q_{ijk}(S_{jk}, S_{jk}) \right) - \mu_k
\]  

(47)

Note that \( \mu_k \) is the unit bandwidth price charged by VP \( k \) to its CBs, and that \( \mu_k \) is the same for all CBs within the VP.

The envelope theorem gives the sensitivity of the aggregate user benefit of a VP with respect to the VP's capacity.

\[
\frac{d}{d V_k} \text{ben}_k(V_k) = \mu_k
\]  

(48)

Equation (48) indicates that the marginal benefit of a virtual path is simply the equilibrium price reached by the VP and its CBs, since no statistical multiplexing is present at this level.

- **Virtual path - Physical trunk Negotiation**

Similarly, the maximization on the physical trunk level is also a concave program. If virtual path \( k \) can obtain the sensitivity of the benefit function, e.g. by (48), it can maximize its consumer surplus by choosing the best virtual path bandwidth allocation \( V_k \) given a price equal to the sum of prices charged by all trunks which the virtual path uses:

\[
\max_{V_k} \text{ben}_k(V_k) - \sum_i \rho_i V_k \quad \text{subject to } V_k \geq 0
\]  

(49)

The optimal solution satisfies

\[
\frac{d}{d V_k} \text{ben}_k(V_k) - \sum_i \rho_i V_k \leq 0
\]  

(50)

\[
V_k \left( \frac{d}{d V_k} \text{ben}_k(V_k) - \sum_i \rho_i V_k \right) = 0
\]  

(51)

Finally on the trunk level, the network performs:

\[
\max_{V_k} \sum_k \text{ben}_k(V_k)
\]  

subject to constraints:

\[
\sum_k V_k I_k \geq T_i
\]  

(52)

\[
V_k \geq 0
\]  

(53)

The Lagrangian function is:

\[
L_i(V_k, \rho_i) = \sum_k \text{ben}_k(V_k) + \rho_i \left( T_i - \sum_k V_k I_k \right)
\]  

where \( \rho_i \geq 0 \)

(54)

Again, trunk \( i \) obtains (55)-(56), virtual path \( k \) obtains (50)-(51).

\[
T_i - \sum_k V_k I_k \geq 0
\]  

(55)

\[
\rho_i \left( T_i - \sum_k V_k I_k \right) = 0
\]  

(56)

Substituting the sensitivity of each VP's benefit given by (48) into (46) and (47), we get

\[
\mu_k - \sum_i \rho_i I_k \leq 0
\]  

(57)

\[
V_k \left( \mu_k - \sum_i \rho_i I_k \right) = 0
\]  

(58)

Equations (58)-(59) say that if the equilibrium price charged by VP \( k \) can not be equal to the sum of the prices charged by all the trunks along the virtual path, then no capacity is allocated to the VP.
**Theorem 3:**

If every level of negotiation in the distributed model converges to its optimal solution, then a globally optimal solution in the centralized model is also achieved.

**Proof:**

Equations (18), (19), (26), (27), (43)-(46), and (56)-(59) derived from the distributed approach are identical to the optimal conditions given by the centralized network maximization problem in Section 3. Q.E.D

4 Identical On-off Markov Fluid Sources

In this section, we would like to study how multiplexing gain, prices and costs respond to changes on the circuit bundle level in the distributed system. We consider a single circuit bundle consisting of identical on-off fluid sources with identical benefit functions.

Since delay jitter \( (D) \) and loss probability are fixed for each circuit bundle in our model, bandwidth and buffer are allocated to a circuit bundle in the same proportion. Figure 1 illustrates the bandwidth and buffer allocated to \( n \) users.

Since users are identical, each is allocated \( E_n = \frac{S_n}{n} \) capacity and \( \frac{B_n}{n} \) buffer, where \( S_n \) is the total CB capacity and \( B_n \) the total CB buffer with \( n \) users respectively.

![Effective bandwidth versus buffer size](image)

**Figure 1** Effective bandwidth versus buffer size

**Theorem 4:**

Buffer \( B_n \) and bandwidth \( S_n \) of a circuit bundle are concave and increasing in the number of users \( n \) in the circuit bundle for on-off Markov fluid sources.

The cost to the user, however, is \( \lambda E_n \), so we must also consider the variation of \( \lambda \) with \( n \). Equation (47) means that for a fixed VP price per effective bandwidth unit \( \mu_k \), a higher marginal aggregate multiplexing gain (MAMG) \(- \sum_i \frac{\partial}{\partial S_{jk}} \mu_k (Q_{ij}^k, S_{jk})\) of a CB within a virtual path lowers the CB's price charged to its VCs. As a result, prices charged by the CBs to their VCs vary within a virtual path, depending on the MAMG of each circuit bundle. For non-real-time service where multiplexing gain does not exist, equation (47) simplifies to \( \lambda_{jk} = \mu_k \).

Adapting equation (47) to the identical-user case, we obtain equation (60).

\[
\lambda_{jk} = \frac{\mu_k}{1 - \sum_i \frac{\partial E_{ij}}{\partial S_{jk}}} = \frac{\mu_k}{1 - n \frac{\partial}{\partial S_n} E(B)} = \frac{\mu_k}{1 - n \frac{\partial}{\partial B} E(B)}
\]

To characterize how the CB price moves with the number of users in the CB, we consider the variation of the MAMG. If no multiplexing gain exists or if demand is less than the channel capacity, then the MAMG is zero. Otherwise MAMG is positive and determined by each individual user's marginal multiplexing gain \( \frac{\partial E_n}{\partial B} \) and by the number of users. These two factors usually work in opposite directions as shown in Figure 1. An increase in the number of users sharing a circuit bundle results in an increase in the number of terms in MAMG. However, it also results in an increase in the channel capacity, \( S_n \), and since effective bandwidth is convex in capacity, this decreases each individual user's marginal multiplexing gain. Sidhu [11] proved that the constant MAMG curve is convex and increasing for the on-off Markovian fluid model. Therefore, if the constant MAMG curve intersects with the constant delay line, the MAMG first increases when the effect of increasing users is dominant, and the MAMG then decreases when the effect of decreasing individual user's multiplexing gain becomes dominant. This relationship for an on-off Markovian fluid source is illustrated in Figure 2a.

The CB price, \( \lambda \), changes inversely to its MAMG (Figure 2b). From a user's point of view, a user values each bandwidth unit more in a CB with higher MAMG. From the network point of view, the network encourages network efficiency by lowering its prices to circuit bundles with higher MAMG, and consequently allows more resources to be allocated to such CBs and more multiplexing gains to be achieved.

In Figure 2c), the cost decreases despite the fact that the price goes up and down.

**Theorem 5:**

For identical on-off Markov fluid sources, cost per user \( C_n \) is convex and decreasing in the number of users \( n \) if the capacity \( S_n \) of a circuit bundle is concave in \( n \).

---

1. State 1’s rate is 64000 and state 2’s rate is 32000. Transition rate from state 1 to 2 is 0.985222 and from 2 to 1 is 0.722282. Maximal delay jitter is 0.05 and the loss probability is 10⁻⁶.
This framework only serves as a primitive starting point for a complete connection establishment process. We focussed on the interplays among resource utilization, user demand and the economic efficiency. To implement a pricing scheme in a real network system might require that it be integrated with other pricing mechanisms addressing different concerns, such as cost recovery and competition. In addition, a richer set of QoS descriptions and corresponding user benefit functions need to be developed. Benefit functions must be dynamically obtained for each network level and the iterative processes must be designed. The convergence of such procedures and their dynamics should be studied. The model can be improved to include the smoothing effect of buffers along a path, which might render extra capacity at the downstream links of a virtual path, and descriptions of user cross elasticity among possible circuit bundles. Furthermore, methods to encourage users to tell the truth about their parameters could be integrated. These problems are discussed in more detail by Jordan [7].

References