Title

Role of price in the connection establishment process

Journal
European transactions on telecommunications and related technologies, 6(4)

ISSN
1120-3862

Authors
Jiang, H
Jordan, S

Publication Date
1995-07-01

Peer reviewed

European Transactions on Telecommunications, vol. 6 no. 4, July-August 1995, pp. 421-429

*The Role of Price in the Connection Establishment Process**

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

hcs@eecs.nwu.edu & scott@eecs.nwu.edu

* This work was supported by the Ameritech foundation.

Abstract

The evolving view of connection establishment for connection-oriented services involves two stages. The first stage consists of separate roles for the user and the network. The user agent must characterize the information streams that will be transmitted and her valuation of the service. Similarly, the network agent must determine the network's resources and its capabilities to accommodate various mixes of service types. The second stage involves negotiations between multiple network and user agents, in which the parties agree to set up connections to transmit the agreed information streams in a manner to guarantee the agreed QoS, and at agreed prices.

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.

I. Introduction

Emerging networks such as Asynchronous Transfer Mode (ATM) will attempt to provide guaranteed performance to variable bit rate (VBR) services. Data networks have generally provided VBR services, but only on a best effort basis. Telephone networks have generally provided guaranteed performance, but only to circuit-switched services.

In ATM, user information streams are organized into small fixed-length packets called *cells*. These cells are sent through shared transmission lines and routed by switches with shared buffers. This resource sharing through cell-switching increases network utilization levels and gives the network the flexibility of providing potentially unlimited classes of service. However, due to the cell-switching nature of ATM, a more complex connection establishment procedure is required to guarantee the quality of services (QoS) and to avoid congestion. Equally important is the economic aspect of connection establishment. It is thought that the decision to accept a call and the amount of resources to assign to it should be based both on the network's capability to accept the call and the network's economic efficiency.

The evolving view of connection establishment for connection-oriented services involves two stages. The first stage consists of separate roles for the user and the network. The user agent must characterize the information streams that will be transmitted and her valuation of the service. Similarly, the network agent must determine the network's resources and its capabilities to accommodate various mixes of service types. The second stage involves negotiations between multiple network and user agents, in which the parties agree to set up connections to transmit the agreed information streams in a manner to guarantee the agreed QoS, and at agreed prices. In [9], 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. 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.

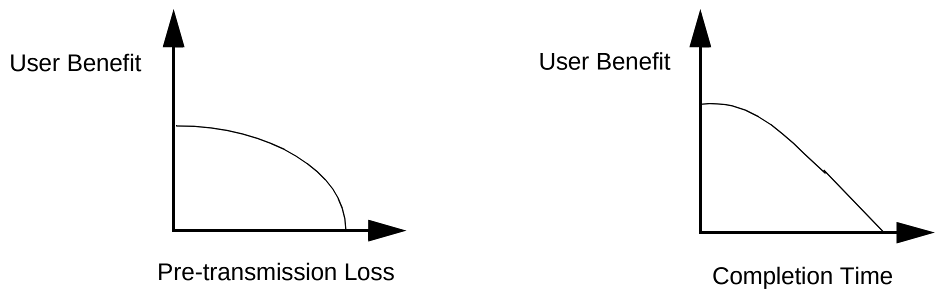
Some recent research has investigated the use of pricing in implementing a distributed contract negotiation process. Best-effort service computer networks have been studied in [1] [2] [3], and ATM networks have been considered in [4] [5] [6] [7] [8] [9]. Although pricing methods for best-effort service 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 [7] emphasized issues of resources allocation among different network levels.

Our pricing scheme in this paper is based on the framework laid in [9], 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 II, the user's role in the first stage of connection establishment is described. User traffic streams are characterized by effective bandwidth and user's valuation of service is defined by benefit functions. Section III outlines the network's role in the first stage by explicitly characterizing the network's capabilities to offer a particular service mix according to the partition of resources at three levels: virtual path, circuit bundle, and virtual circuit. In section IV, the structure and the frequencies of the negotiation processes are explained. In sections V and VI, mathematical models are established based on the framework described in previous sections. Finally, in the last section, we discuss remaining issues concerning design and convergence of iterative negotiation processes.

II. User Characterization

We assume a user's objective is to maximize her consumer surplus, defined as the difference between benefit and cost. We model a user's demand for bandwidth *for a real-time service* through user benefit as a function of pre-transmission loss (Figure 1a). For example, a user might adjust the compression levels for her video transmission according to prices for bandwidth.



a) Benefit function for real-time service

b) Benefit function for non-real-time services

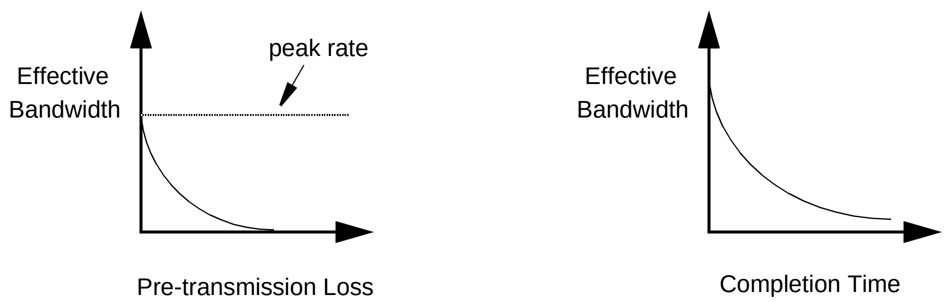
Figure 1 User benefit versus user demand elasticity

We model a user's demand for bandwidth f*or a non-real-time service* through user benefit as a function of completion time (Figure 1b). 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 change with elapsed time and remaining file size.

We choose effective bandwidth to characterize user traffic streams. Effective bandwidth [10] [11] [12] [13] [14] [15] models a source's use of network resources at the user/network interface. A source is fed into a finite buffer served at a constant rate. If the loss is asymptotically exponential in the buffer length, then the source is said to have an *effective bandwidth*. The rate of the exponential decrease depends on the service rate and on the burstiness of the source. The concept is usually used in reverse: in the range of small loss probabilities and large buffer lengths, a source must be served at a rate at least equal to its effective bandwidth in order to meet the corresponding loss criterion.

The effective bandwidth is bounded by the source's mean and peak rates, and is a function of the source's burstiness and of $\xi = (\log(\text{lossprobability}))/(\text{bufferlength})$. One nice property of many systems with effective bandwidths is that the effective bandwidth of multiplexed sources sharing the same buffer is equal to the sum of their individual effective bandwidths and hence only depends on the characterization of individual traffic sources and the parameter ξ. Effective bandwidth thus serves as a useful tool in allocating network resources to satisfy a source's QoS.

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, as depicted in Figure 2, to characterize user traffic streams.



a) Effective bandwidth for real-time services b) Effective bandwidth for non-real-time services

Figure 2 Effective bandwidth versus demand elasticity

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 (Figure 2b).

III. 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]. The network's objective can also be to maximize the economic efficiency in terms of revenue [17] [18] [19] or a combination of revenue and user benefit, depending on cost, regulation, market competition, etc. 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 this 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. [20] [21]).

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) [16]. A virtual path is a group of connections sharing a common path from source to destination. Virtual circuits are the individual connections within a virtual path. Routing is performed on the virtual path level. A circuit bundle groups VCs with common characteristics as given by the resource management architecture. Among the numerous combinations, the following allocation methods are often considered:

Circuit Switching: Each VC is allocated its own bandwidth and buffers.

VP/QoS Allocation: All VCs with identical paths and QoS are statistically multiplexed.

Effective bandwidth is used as a tool for admission control. Under a circuit switching architecture, each virtual circuit is allocated the amount of bandwidth equal to its effective bandwidth, given the traffic and QoS specifications. 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.

IV. Contract Negotiation

Both centralized and distributed approaches can be designed to match user needs with network availability. In a centralized approach, the network determines the most efficient resource allocation by collecting complete user traffic characterizations and user valuations of each service, and then enforces its decision on individual users and on network controls at various network levels. We consider this approach in sections V.A and section VI. The drawbacks to a centralized approach are heavy traffic to and a large computation burden on the central node. We consider an iterative and distributed negotiation process through pricing in section V.B.

Prices play an important role in implementing the distributed negotiation processes. By setting prices, the network agent signals to the user agents the available capacity and the market demand. A user agent reacts to the price by choosing the amount of resources to demand. At market-clearing prices, the network and user agents agree on the prices and the amount of resources for each connection.

We argue that effective-bandwidth pricing reflects an individual user's usage of the network, and thus is better than peak-rate pricing or mean-rate pricing when statistical multiplexing is considered in admission control. Peak-rate pricing charges more than what a VBR traffic source actually uses due to the multiplexing gain. Mean-rate pricing, or per-packet pricing, does not reflect the cost caused by the burstiness of traffic sources. 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: $\text{meanrate} + \text{burstiness}$. We therefore propose that users be charged an amount equal to *effective bandwidth x price*. 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 each level communicates only with its adjacent levels. Each circuit bundle, virtual path and trunk maximizes its total user benefit as a supplier as it negotiates with lower network levels by setting the right prices; and maximizes its consumer surplus as a consumer as it negotiates with the 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. Three levels of negotiations thus take place on different time scales with the lowest level having the shortest negotiation cycle. Consequently, computations of optimal resource allocations are distributed among network levels and among local network areas.

V. VP/QoS Allocation

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 adopt the VP/QoS allocation architecture for resource management, where virtual circuits sharing the same virtual path and QoS are grouped into a circuit bundle. 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 this resource allocation scheme, virtual circuits within a virtual path are grouped into different bundles (circuit bundles) according to their QoS[1] specifications. Virtual circuits within a circuit bundle are statistically multiplexed, 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_{ijk}: effective bandwidth of virtual circuit (VC) i within circuit bundle (CB) j which is within virtual path (VP) k.

S_{jk}: capacity of CB j within VP k.

T_l: capacity of trunk l.

V_k: capacity of VP k.

Q_{ijk}: pre-transmission loss for real-time services and completion time for non-real-time services.

ben_{ijk}: benefit function for VC i within CB j and within VP k, or benefit function of user (i, j, k).

$ben_{jk} = \sum_i ben_{ijk}$: aggregate benefit function of circuit bundle (j, k).

$ben_k = \sum_j \sum_i ben_{ijk}$: aggregate benefit function of virtual path k.

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

I: a routing matrix with I_{kl} as elements.

λ_{jk}: the Lagrangian multiplier and the unit price in \$/bit/second of CB j within VP k.

μ_k: the Lagrangian multiplier and the unit price in \$/bit/second of VP k.

ρ_l: the Lagrangian multiplier and the unit price in \$/bit/second of trunk l.

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 Q_{ijk}. Higher pre-transmission loss results in less traffic being transmitted onto the network, and therefore 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.

---

1. QoS of the circuit bundle here is defined by cell loss probability and maximal delay jitter.

The relationship between effective bandwidth and channel capacity for a 2 state Markov modulated fluid flow[1] is shown in Figure 3.

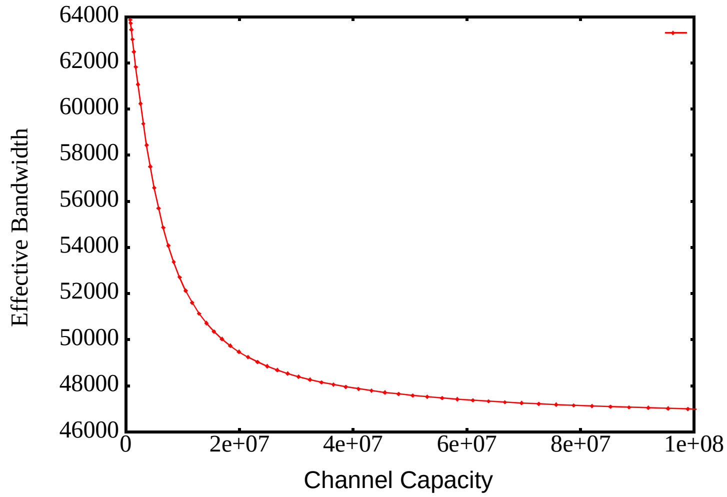


Figure 3 Effective bandwidth versus channel capacity for fixed QoS

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}.

A. Centralized Network Maximization

Assuming that the network knows its trunk capacities and virtual path routing, and every user's benefit function and traffic stream characterization, the network will be able to calculate the optimal resource allocation for different network levels. The network performs total user benefit maximization for fixed routing and trunk sizes:

$$\max_{Q_{ijk}, S_{jk}, V_k} \sum_i \sum_j \sum_k ben_{ijk}(Q_{ijk}) \quad (1)$$

subject to constraints:

$$\sum_i E_{ijk}(Q_{ijk}, S_{jk}) \le S_{jk} \qquad \forall(j, k) \quad (2)$$

$$\sum_j S_{jk} \le V_k \qquad \forall k \quad (3)$$

1. The parameters of the fluid flow are: fluid rates (64000,32000), transition rates (.985222,.722282). The QoS is defined by a maximal delay jitter of 0.05 and a loss probability of 10^{-9}.

$$\sum_k V_k I_{kl} \le T_l \qquad \forall l \tag{4}$$

$$Q_{ijk}, S_{jk}, V_k \ge 0 \qquad \forall (i, j, k) \tag{5}$$

The Lagrangian function is:

$$L(\lambda_{jk}, \mu_k, \rho_l) = \sum_i \sum_j \sum_k ben_{ijk}(Q_{ijk}) + \sum_j \sum_k \lambda_{jk}\left(S_{jk} - \sum_i E_{ijk}\right) + \sum_k \mu_k\left(V_k - \sum_j S_{jk}\right) + \sum_l \rho_l\left(T_l - \sum_k V_k I_{kl}\right) \tag{6}$$

$S_{jk} - \sum_i E_{ijk}(Q_{ijk}, S_{jk})$ is a concave function since $\sum_i E_{ijk}(Q_{ijk}, S_{jk})$ is convex. Because the objective function and the constraint functions in (6) are concave, this is a concave programming problem and Kuhn-Tuker conditions [22] are sufficient and necessary for the global optimal solution. Kuhn-Tuker theorem gives (2)-(4) plus the following conditions for optimality:

$$\frac{\partial}{\partial Q_{ijk}} L(\lambda_{jk}, \mu_k, \rho_l) = \frac{\partial}{\partial Q_{ijk}} ben_{ijk}(Q_{ijk}) - \lambda_{jk} \frac{\partial E_{ijk}}{\partial Q_{ijk}} \le 0 \qquad \forall (i, j, k) \tag{7}$$

$$Q_{ijk}\left(\frac{\partial}{\partial Q_{ijk}} ben_{ijk}(Q_{ijk}) - \lambda_{jk} \frac{\partial E_{ijk}}{\partial Q_{ijk}}\right) = 0 \qquad \forall (i, j, k) \tag{8}$$

$$\frac{\partial}{\partial S_{jk}} L(\lambda_{jk}, \mu_k, \rho_l) = \lambda_{jk} - \lambda_{jk} \sum_i \frac{\partial E_{ijk}}{\partial S_{jk}} - \mu_k \le 0 \qquad \forall (j, k) \tag{9}$$

$$S_{jk}\left(\lambda_{jk} - \lambda_{jk} \sum_i \frac{\partial E_{ijk}}{\partial S_{jk}} - \mu_k\right) = 0 \qquad \forall (j, k) \tag{10}$$

$$\frac{\partial}{\partial V_k} L(\lambda_{jk}, \mu_k, \rho_l) = \mu_k - \sum_l \rho_l I_{kl} \le 0 \qquad \forall l \tag{11}$$

$$V_k\left(\mu_k - \sum_l \rho_l I_{kl}\right) = 0 \qquad \forall l \tag{12}$$

$$\lambda_{jk}\left(S_{jk} - \sum_i E_{ijk}\right) = 0 \qquad \forall (j, k) \tag{13}$$

$$\mu_k\left(V_k - \sum_j S_{jk}\right) = 0 \qquad \forall k \tag{14}$$

$$\rho_l\left(T_l - \sum_k V_K I_{kl}\right) = 0 \qquad \forall l \tag{15}$$

$$\lambda_{jk}, \mu_k, \rho_l \ge 0 \qquad \forall (j, k, l) \tag{16}$$

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 λ_{jk} is the real price charged to its users. The VP price μ_k and trunk price ρ_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.

B. Distributed Network Maximization

Network maximization is distributed to and performed at three network levels: user - circuit bundle, circuit bundle - virtual path, virtual path - physical trunk. At each level, an equilibrium price is reached and total user benefit is locally maximized.

• User - Circuit bundle Negotiation

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

$$\max_{Q_{ijk}} \; ben_{ijk}(Q_{ijk}) - \lambda_{jk}E_{ijk}(Q_{ijk}, S_{jk}), \quad \text{subject to} \, Q_{ijk} \geq 0 \tag{17}$$

The optimal solution satisfies:

$$\frac{\partial}{\partial Q_{ijk}} ben_{ijk}(Q_{ijk}) - \lambda_{jk}\frac{\partial E_{ijk}}{\partial Q_{ijk}} \leq 0 \tag{18}$$

$$Q_{ijk}\left(\frac{\partial}{\partial Q_{ijk}} ben_{ijk}(Q_{ijk}) - \lambda_{jk}\frac{\partial E_{ijk}}{\partial Q_{ijk}}\right) = 0 \tag{19}$$

(18) and (19) show that if the marginal benefit ($-\frac{\partial}{\partial Q_{ijk}} ben_{ijk}(Q_{ijk})$) is strictly greater than the marginal cost of effective bandwidth ($-\lambda_{jk}\frac{\partial E_{ijk}}{\partial Q_{ijk}}$) 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:

$$\frac{\partial}{\partial Q_{ijk}} ben_{ijk}(Q_{ijk}) - \lambda_{jk}\frac{\partial E_{ijk}}{\partial Q_{ijk}} = 0 \tag{20}$$

(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:

$$\max_{Q_{ijk}} \sum_i ben_{ijk}(Q_{ijk}) \tag{21}$$

subject to constraints:

$$\sum_i E_{ijk}(Q_{ijk}, S_{jk}) \le S_{jk} \tag{22}$$

$$Q_{ijk} \ge 0 \tag{23}$$

The Lagrangian function is:

$$L_{jk}(Q_{ijk}, \lambda_{jk}) = \sum_i ben_{ijk}(Q_{ijk}) + \lambda_{jk}\left(S_{jk} - \sum_i E_{ikj}\right), \quad \text{where } \lambda_{jk} \ge 0 \tag{24}$$

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

$$L_{jk}(Q'_{ijk}, \lambda'_{jk}) = \max_{Q_{ijk}} \min_{\lambda_{jk}} L_{jk}(Q_{ijk}, \lambda_{jk}) \tag{25}$$

(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 λ_{jk}, which leads to

$$S_{jk} - \sum_i E_{ijk}(Q_{ijk}, S_{jk}) \ge 0 \tag{26}$$

$$\lambda_{jk}\left(S_{jk} - \sum_i E_{ijk}(Q_{ijk}, S_{jk})\right) = 0 \tag{27}$$

On the other hand, user (i, j, k) maximizes L_{jk}, for a fixed circuit bundle price λ_{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 such as that described in section IV.

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

$$\frac{d}{dS_{jk}} ben_{jk}(S_{jk}) = \left.\frac{\partial L_{jk}}{\partial S_{jk}}\right|_{Q_{ijk} = Q'_{ijk}(S_{jk}),\, \lambda_{jk} = \lambda'_{jk}(S_{jk})} = \lambda'_{jk}(S_{jk})\left(1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk}(Q'_{ijk}(S_{jk}), S_{jk})\right) \tag{28}$$

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

The marginal benefit of a circuit bundle depends the equilibrium price reached between the CB and its VCs, λ'_{jk}, and on the marginal aggregate multiplexing gain (MAMG) within the CB, $-\sum_i \frac{\partial E'_{ikj}}{\partial S_{jk}}$. 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 and by the number of users. These two factors usually work in opposite directions. 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_{jk}, and since effective bandwidth is convex in capacity, this decreases each individual user's marginal multiplexing gain. Therefore it might be possible that MAMG first increases when the effect of increasing users is dominant, and then decreases when the effect of decreasing individual user multiplexing gain becomes dominant. This relationship (for the source considered in Figure 3) is shown in Figure 4.

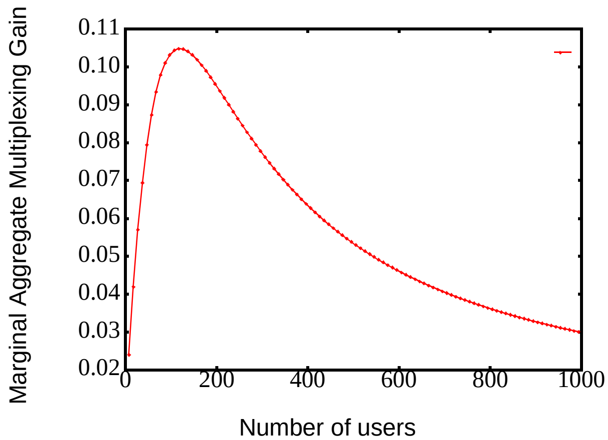


Figure 4 Marginal aggregate multiplexing gain versus number of users

- **Circuit bundle - Virtual path Negotiation**

If we can obtain the sensitivity of the benefit function $ben_{jk}(S_{jk})$, e.g. by (28), then the circuit bundle is able to maximize its consumer surplus given its VP's price μ_k. Circuit bundle (j, k) performs:

$$\max_{S_{jk}} \; ben_{jk}(S_{jk}) - \mu_k S_{jk}, \quad \text{subject to} S_{jk} \geq 0 \tag{29}$$

The optimal solution satisfies

$$\frac{\partial}{\partial S_{jk}} ben_{jk}(S_{jk}) - \mu_k \leq 0 \tag{30}$$

$$S_{jk}\left(\frac{\partial}{\partial S_{jk}} ben_{jk}(S_{jk}) - \mu_k\right) = 0 \tag{31}$$

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

$$\max_{S_{jk}} \; \sum_j ben_{jk}(S_{jk}) \tag{32}$$

subject to constraints:

$$\sum_j S_{jk} \leq V_k \tag{33}$$

$$S_{jk} \geq 0 \tag{34}$$

The Lagrangian function is:

$$L_k(S_{jk}, \mu_k) = \sum_j ben_{jk}(S_{jk}) + \mu_k\left(V_k - \sum_j S_{jk}\right), \quad \text{where} \quad \mu_k \geq 0 \tag{35}$$

Using the same saddle-value argument, virtual path k obtains (36)-(37), and circuit bundle (j, k) obtains (30)-(31), 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.

$$\mu_k\left(V_k - \sum_j S_{jk}\right) = 0 \qquad \forall k \tag{36}$$

$$V_k - \sum_j S_{jk} \geq 0 \qquad \forall k \tag{37}$$

Substituting the sensitivity of each bundle's benefit to allocated bandwidth given by (28) into (39) and (40), we get

$$\lambda'_{jk}\left(1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk}(Q'_{ijk}(S_{jk}), S_{jk})\right) - \mu_k \leq 0 \qquad \forall j \tag{38}$$

$$S_{jk}\left(\lambda'_{jk}\left(1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk}(Q'_{ijk}(S_{jk}), S_{jk})\right) - \mu_k\right) = 0 \qquad \forall j \tag{39}$$

This complementary slackness condition indicates that if the second term in (39) is less than zero, then no capacity will be allocated to CB (j, k). If the second term is equal to zero, then (40) has to be true.

$$\lambda'_{jk}\left(1 - \sum_i \frac{\partial}{\partial S_{jk}} E_{ijk}(Q'_{ijk}, S_{jk})\right) = \mu_k \tag{40}$$

Note that μ_k is the unit bandwidth price charged by VP k to its CBs, and thus μ_k is the same for all CBs within the VP. Equation (40) means that for a fixed VP price per effective bandwidth unit μ_k, a higher marginal aggregate multiplexing gain (MAMG) $-\sum_i \frac{\partial}{\partial S_{jk}} E_{ijk}(Q'_{ijk}, S_{jk})$ 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 marginal aggregate multiplexing gain of each circuit bundle. For non-real-time service where multiplexing gain does not exist, equation (40) simplifies to $\lambda'_{jk} = \mu_k$.

Intuitively, 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 circuit bundles and more multiplexing gains to be achieved.

The envelope theorem gives

$$\frac{d}{dV_k} ben_k(V_k) = \mu'_k(V_k) \tag{41}$$

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

If virtual circuit k can obtain the sensitivity of the benefit function, e.g. by (41), 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:

$$\max_{V_k} \; ben_k(V_k) - V_k \sum_l \rho_l I_{kl}, \quad \text{subject to} V_k \geq 0 \tag{42}$$

The optimal solution satisfies

$$\frac{\partial}{\partial V_k} ben_k(V_k) - \sum_l \rho_l I_{kl} \leq 0 \tag{43}$$

$$V_k \left(\frac{\partial}{\partial V_k} ben_k(V_k) - \sum_l \rho_l I_{kl} \right) = 0 \tag{44}$$

Finally on the trunk level, the network performs:

$$\max_{V_k} \; \sum_k ben_k(V_k) \tag{45}$$

subject to constraints:

$$\sum_k V_k I_{kl} \leq T_l \tag{46}$$

$$V_k \geq 0 \tag{47}$$

The Lagrangian function is:

$$L_l(V_k, \rho_l) = \sum_k ben_k(V_k) + \rho_l \left(T_l - \sum_k V_k I_{kl} \right), \text{ where } \rho_l \geq 0 \tag{48}$$

Again, trunk l obtains (49)-(50), virtual path k obtains (43)-(44).

$$T_l - \sum_k V_k I_{kl} \geq 0 \tag{49}$$

$$\rho_l \left(T_l - \sum_k V_k I_{kl} \right) = 0 \tag{50}$$

Substituting the sensitivity of each VP's benefit given by (41) into (49) and (50), we get

$$\mu'_k - \sum_l \rho_l I_{kl} \leq 0 \tag{51}$$

$$V_k \left(\mu'_k - \sum_l \rho_l I_{kl} \right) = 0 \tag{52}$$

Equations (51)-(52) 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.

Equations (18), (19), (26), (27), (36)-(39), and (49)-(52) derived from the distributed approach are identical to the optimal conditions given by the centralized network maximization problem in section V.A. This means if every negotiation level converges to its optimal point, then an globally optimal point is achieved.

VI. Circuit Switching Allocation

Circuit switching is the simplest form of resource allocation, where each virtual circuit is allocated its own bandwidth and buffer according to its QoS specification. Consequently, no statistical multiplexing exists in this allocation form. In Section II, we were able to characterize user benefit as a function of effective bandwidth. In essence, we used effective bandwidth as a measure of grade of service which includes the QoS of the network channel and user demand elasticity. Since effective bandwidth is uniquely determined by Q_{ijk} and user traffic characterization in absence of statistical multiplexing, $ben_{ijk}(Q_{ijk})$ is equivalent to $ben_{ijk}(E_{ijk})$.

The network performs:

$$\max_{E_{ijk}, S_{jk}, V_k} \sum_i \sum_j \sum_k ben_{ijk}(E_{ijk}) \tag{53}$$

subject to constraints:

$$\sum_i E_{ijk} \le S_{jk} \qquad \forall (j, k) \tag{54}$$

$$\sum_j S_{jk} \le V_k \qquad \forall k \tag{55}$$

$$\sum_k V_k I_{kl} \le T_l \qquad \forall l \tag{56}$$

$$E_{ijk}, S_{jk}, V_k \ge 0 \qquad \forall (i, j, k) \tag{57}$$

The Lagrangian function of the above maximization problem is:

$$L(\lambda_{jk}, \mu_k, \rho_l) = \sum_i \sum_j \sum_k ben_{ijk}(E_{ijk}) + \sum_j \sum_k \lambda_{jk}\left(S_{jk} - \sum_i E_{ikj}\right) + \sum_k \mu_k\left(V_k - \sum_j S_{jk}\right) + \sum_l \rho_l\left(T_l - \sum_k V_k I_{kl}\right) \tag{58}$$

Kuhn-Tuker theory gives the following conditions including (54)-(56) that an optimal solution satisfies:

$$\frac{\partial}{\partial E_{ijk}} L(\lambda_{jk}, \mu_k, \rho_l) = \frac{\partial}{\partial E_{ijk}} ben_{ijk}(E_{ijk}) - \lambda_{jk} \le 0 \qquad \forall (i, j, k) \tag{59}$$

$$E_{ijk}\left(\frac{\partial}{\partial E_{ijk}} ben_{ijk}(E_{ijk}) - \lambda_{jk}\right) = 0 \qquad \forall (i, j, k) \tag{60}$$

$$\frac{\partial}{\partial S_{jk}} L(\lambda_{jk}, \mu_k, \rho_l) = \lambda_{jk} - \mu_k \le 0 \qquad \forall (j, k) \tag{61}$$

$$S_{jk}(\lambda_{jk} - \mu_k) = 0 \qquad \forall(j, k) \tag{62}$$

$$\frac{\partial}{\partial V_k} L(\lambda_{jk}, \mu_k, \rho_l) = \mu_k - \sum_l \rho_l I_{kl} \le 0 \qquad \forall k \tag{63}$$

$$V_k\left(\mu_k - \sum_l \rho_l I_{kl}\right) = 0 \qquad \forall k \tag{64}$$

$$\lambda_{jk}\left(S_{jk} - \sum_i E_{ijk}\right) = 0 \qquad \forall(j, k) \tag{65}$$

$$\mu_k\left(V_k - \sum_j S_{jk}\right) = 0 \qquad \forall k \tag{66}$$

$$\rho_l\left(T_l - \sum_k V_K I_{kl}\right) = 0 \qquad \forall l \tag{67}$$

$$\lambda_{jk}, \mu_k, \rho_l \ge 0 \qquad \forall(j, k, l) \tag{68}$$

(61) and (62) means CB price and VP price are equal for the circuit bundles having positive capacity allocations from the same virtual path. Therefore, under this network architecture, the circuit bundle level virtually disappears since this level does not differentiate itself from other network levels either in admission control or pricing.

This centralized problem can also be decomposed into a distributed problem as we did in the last subsection. In general, this problem is a simplified version of the last problem. Total user benefit of the network is maximized when the equilibrium prices, obtained through the distributed negotiation process, equal the Lagrangian multipliers $\lambda_{jk}, \mu_k, \rho_l$, where λ_{jk} is the optimal price charged by circuit bundle (i, j) to its virtual circuits, μ_k is the price charged by virtual path k to its circuit bundles, and ρ_l is the price charged by trunk l to each virtual path going through the trunk.

VII. Parting Thoughts

We developed a framework for a complete connection establishment procedure to achieve network efficiency through distributed, iterative and hierarchical negotiation processes. Pricing is used as an incentive mechanism to carry out the negotiation processes and to signal the optimality of resource allocation. Except for the cases where demand is less than supply, the optimal prices charged by circuit bundles are equal to the optimal price by their virtual path multiplied by a factor associated with multiplexing gain. In case of non-real-time service or circuit switching, the optimal CB prices and their VP price are equal. The optimal price charged by a virtual path is equal to the sum of the optimal prices by the physical trunks it passes.

Statistical multiplexing is incorporated in the pricing model and the effects of multiplexing gain on prices and resource allocation are studied. In summary, higher marginal aggregate multiplexing gain of a circuit bundle lowers the price charged by the circuit bundle to its users and allows the circuit bundle to obtain more capacity from its virtual path than other circuit bundles within the virtual path.

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 in Jordan [9].

References

[1] R. Cocchi, D. Estrin, S. Shenker and L. Zhang. “Pricing in computer networks: motivation, formulation, and example.” *IEEE/ACM Transaction on Networking*, 1(6):614-627,December 1993.

[2] J. K. Mackie-Mason and H. R. Varian. “Pricing the Internet”, *Second International Conference on Telecommunication Systems Modelling and Analysis*, pp378-393, Nashville, Tennessee, March 24-27, 1994.

[3] C. Parris, S. Keshav and D. Ferrari. “A framework for the study of pricing in integrated networks.” *Technical Report TR-92-016*, International Computer Science Institute, Berkeley, CA, March 1992.

[4] F. P. Kelly. “On tariffs, policing and admission control for multiservice networks.” *Operations Research Letters*, 15:1-9, 1994.

[5] S. H. Low and P. P. Varaiya. “A new approach to service provisioning in ATM networks.” *IEEE Transactions on Networking*, 1(3):547-553, 1883.

[6] J. Murphy and L. Murphy. “Bandwidth allocation by pricing in ATM networks.” preprint.

[7] J. Murphy, L. Murphy and E. C. Posner. “Distributed pricing for embedded ATM networks.” *International Teletraffic Congress ITC-14*, Antibes, France, June 1994.

[8] C. Parris and D. Ferrari. “A resource based pricing policy for real-time channels in a packet-switching network.” preprint.

[9] S. Jordan and H. Jiang. “Connection Establishment in High Speed Networks.” submitted to *IEEE Journal on Selected Areas in Communications*.

[10] J. Y. Hui. “Resource allocation for broadband networks.” *IEEE Journal on Selected Areas in Communications*, 6(9):1598-1608, December 1988.

[11] F. P. Kelly. “Effective bandwidths at multi-class queues.” *Queueing Systems*, 9: 5-15, September, 1991.

[12] R. J. Gibbens and P. J. Hunt. “Effective bandwidths for the multi-type UAS channel.”, *Queueing Systems*, 9:17-28, May 1991.

[13] R. Guerin, H. Ahmadi and M. Naghshineh. “Equivalent capacity and its application to bandwidth allocation in high-speed networks.” *IEEE Journal on Selected Areas in Communications*, 9(7):968-981, September 1991.

[14] G. Kesidis, J. Walrand and C. S. Chang. “Effective bandwidths for multiclass markov fluids and other ATM sources.” *IEEE/ACM Transactions on Networking*, 1(4):424-428, August 1993.

[15] A. I. Elwalid and D. Mitra. “Effective bandwidth of general markovian traffic sources and admission control of high speed networks.” *IEEE/ACM Transactions on Networking*, 1(3):329-343, June 1993.

[16] J. Burgin and D. Dorman. “Broadband ISDN resource management: the role of virtual paths.” *IEEE*

Communications Magazine, 29(9):44-48, September 1993.

[17] S. Jordan and P. P. Varaiya. "Throughput service, multiple resource communication networks." *IEEE Transactions*, 39(5):1216-1222, August 1991.

[18] F. P. Kelly. "Routing in circuit-switched networks: Optimization, shadow prices and decentralization." *Advanced Applied Probability*, 20: 112-144, 1988.

[19] M. Honig and K. Steiglitz. "Usage-based pricing and quality of service in data networks." Submitted to *IEEE Journal on Selected Areas in Communications*.

[20] B. Mitchell and I. Vogelsang. "Telecommunications pricing." *Cambridge University Press*, 1991.

[21] R. P. McLean and W. W. Sharkey. "An approach to the pricing of broadband telecommunications services." To appear in *Telecommunications Systems*.

[22] J. Franklin. "Methods of mathematical economics: linear and nonlinear programming: fixed-point theorems." *New York: Springer-Verlag*, 1980.