An Approach to Manage Connectionless Services in Connection-Oriented Networks

Muhammet ABDELATI and Erdal ARIKAN
Electrical and Electronics Engineering Department, Bilkent University 06533, Ankara, Turkey
Phone (90) 312-2664862 Fax (90) 312-2664126 E-mail: muhammet@ee.bilkent.edu.tr

Abstract - In this work we propose a pricing scheme which serves as an instrument for managing connectionless services in connection-oriented communication networks. The scheme is able to allocate network bandwidth in a Pareto-optimal way that maximizes the total surplus. The key idea is to decompose the service provision procedure among three separate parties whose interactions are governed by a set of competitive pricing mechanisms.

I. INTRODUCTION

Connection-oriented networks are well-suited to handle interactive and real-time applications such as telephony and video conferencing. However, they will be underutilized if used directly in applications characterized by sporadic behavior and short service time requirements such as mail and file transfer. For such applications, in which the user information typically consists of a single block of data, connection-oriented services are inefficient due to connection establishment and tear-down overhead. Such applications are more efficiently handled by a connectionless service which multiplexes data from individual applications into a pre-established virtual connection.

The aim of this paper is to propose a method for providing connectionless services in a connection-oriented network. (An important instance of this problem, which motivated the present work, is LAN interconnection over an ATM network.) Several solutions have been proposed for this problem [1, 2, 3]. A general discussion of this problem can be found in [4, 5, 6].

The approach proposed in this paper is based on a pricing scheme which allocates the communication resources in a Pareto-optimal way that achieves maximum total surplus [7]. In fact, pricing plays an important role in any complete network architecture. If services were free, there would be no incentive to request less than the best service the network could provide, which would not produce effective utilization of the network resources. There have been several papers dealing with pricing as a means of resource allocation in integrated networks, see, e.g., [8, 9, 10]; however, none used it as a means of providing connectionless traffic support in connection-oriented networks.

II. THE PROPOSED SCHEME

In certain aspects, our work falls between the reservation-oriented approach in [9, 10] and the best-effort type approach in [8]. Description of the proposed scheme is given in Section II. Section III presents some simulation results and finally in Section IV, comments and conclusions are given.

A. The Model

For simplicity of exposition, we consider here a simple network consisting of a single origin-destination pair \((o-d)\) connected by a virtual channel of capacity \(K\) channels. Each channel is assumed to be a delayless, errorless bit pipe. An arbiter leases from the network a certain number of channels, which we call a box, for a fixed time period that is orders of magnitude longer than the duration of a typical connectionless service. Periodically, the network holds auctions to serve the entry or expansion requests of the arbiters. The network charges the arbiters a per-unit-time channel fee of \(p(k)\) where \(k\) is the total number of channels hired to all the arbiters.

Fig. 1. Functional decomposition.
Table 1: Traffic Specifications.

<table>
<thead>
<tr>
<th>Class</th>
<th>Example</th>
<th>a (chans)</th>
<th>T (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>E-mail</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>FTP</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>D-base query</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

There are 4 service classes labeled A, B, C and D. Class $x \in \{A, B, C, D\}$ is characterized by a pair $(a_x, T_x)$ indicating that the service requires $a_x$ channels for a duration of $T_x$ as shown in Table 1. Only class A traffic is assumed to be insensitive to queuing delay.

Users belonging to service class $x$ arrive at the network at a rate of $\lambda_x = \lambda_x(s_x)$ where $p_x$ is the prevailing per-unit-time-channel price charged to users of service class $x$. A user of type $x \in \{B, C, D\}$ arriving at time $t$ searches for an arbiter who has $a_x$ channels free; if none exists, the user is lost. If an arbiter exists with $a_x$ free channels, the user is served for $T_x$ seconds and pays the arbiter the amount $p_x T_x$. Users of type A are never lost; they join the queue of a bus and are served whenever the bus is idle. The quality of service is measured by the delay for service class A and the probability of loss for classes B, C, and D.

B. Pricing of Services

We propose a pricing scheme based on risk analysis that is well known in finance literature [11, 12]. Let $\phi_x$ be an index which reflects the randomness of revenues of service class $x$. The specific risk index used here is defined in the next subsection. The arbiters set the price for service $x$ as

$$ p_x = p_x(1 + \rho(\phi_x)) \tag{1} $$

where the risk index and the function $\rho$ are such that $\rho(\phi_x) = 0$. We call $\rho(\phi_x)$ a risk premium charged to class $x$ since the uncertainty in future revenues is a type of risk and arbiters ask a premium as compensation.

In free competition markets there are many arbiters, and users will be attracted to that arbiter which asks the lowest price. Hence, the market price function will be the lower envelope of the price functions offered by arbiters as illustrated in Fig. 2.

Any arbiter who would like to persist in the market should have at least a portion of its price function tangent to the market price function. Competition and free-entry into market will always force arbiters to lower their price functions.

C. Risk Index

A class $x$ user who arrives at time $t_j$ generates a potential per-unit-time revenue $r_{x,j}(t)$ given by

$$ r_{x,j}(t) = \begin{cases} a_x p_x & t_j < t < t_j + T_x \\ 0 & \text{else} \end{cases} \tag{2} $$

Let $R_x(t)$ be a random process representing the revenue of service class $x$ normalized to have unit mean; namely,

$$ R_x(t) = \frac{1}{E(\sum_j r_{x,j}(t))} \sum_j r_{x,j}(t) \tag{3} $$

where $E\{\cdot\}$ denotes expectation. Note that $R_x(t)$ includes the potential revenue unrealized due to unmet demand. The risk index used in this work is

$$ \phi_x(t) = \frac{\text{Cov}(R_x(t), R_m(t))}{\sigma_h(t)} \tag{4} $$

where $R_m(t)$ is the market revenue given by

$$ R_m(t) = \sum_{x \in \{A, B, C, D\}} R_x(t) \tag{5} $$

Note that $\phi_A = 0$ and $\sum_{x \in \{A, B, C, D\}} \phi_x = 1$

D. Equilibrium Analysis

There are several approaches to investigate the welfare properties of the competitive markets. The one we pursue here, the discrete good model approach, is probably the simplest. In this approach there are two goods: $y$ and $z$. $y$ is the channels and $z$ is the money left for purchasing other goods. A user utility function is defined as

$$ u_j(y, z) = u_j(s_j, m_j - a_j p) \tag{6} $$

where $p$ is the channel's price and $m_j$ is the user's budget. The user's surplus, $s_j$, is defined as the change in his utility after making a decision. That is

$$ s_j = u_j(s_j, m_j - a_j p) - u(0, m_j) \tag{7} $$

The reservation price is that price $p_j$ which just makes the consumer indifferent to being served or not. That is, it is the price $p_j$ which satisfies the equation

$$ u_j(s_j, m_j - a_j p_j) = u_j(0, m_j) \tag{8} $$

1695
Assuming quasilinear preferences \([7], (9)\) may be rewritten as
\[
u_j(s_j) = u_j(0) + m_j
\] (9)

Substituting in (7), one can easily show that the user's surplus reduces to
\[
s_j = \begin{cases} 
  u_j(p_j - p) & p < p_j \\
  0 & \text{else} 
\end{cases} 
\] (10)

The aggregate users' surplus, \(S_u\), is the sum of individual surpluses. To find the network surplus at an output level of \(K\) channels, note that all \(K\) channels are hired with a price of \(p(K)\). However, the \(k\)th channel \((k = 1, \ldots, K)\) has an opportunity cost of \(p(k)\). Therefore, it contributes to the network surplus the amount \(p(K) - p(k)\), and the total network surplus is given by
\[
S_n = \sum_{k=1}^{K} [p(K) - p(k)],
\] (11)

The arbiters' surplus is defined to be their total profits given by
\[
S_p = K(p - p(K)).
\] (12)

Defining a social welfare function \(W = S_u + S_n + S_v\), one can show that the competitive equilibrium \((p^*, K^*)\) is Pareto-optimal and maximizes \(W\) \([7]\).

### III. Simulation Results

In this section we present simulation results for an example that demonstrates the basic features of our framework. We compare our competitive market with a monopolistic one. The example network consists of a single node and an output link of capacity 75 M B/s which is equivalent to 750 channels. We assume that it is desired to allocate not more than 25% of the link capacity to connectionless traffic. To this end, the network uses the price function
\[
p(k) = \frac{3000}{750 - 4k},
\] (13)

where \(k\) is the total number of channels hired. This function is rather simple and satisfies our objectives of increasing slowly when resources are abundant \((\eta = k/750 < 0.1)\) while increasing rapidly when they become scarce \((\eta > 0.2)\).

In the simulation the arrival process for service class \(z\) is Poisson with rate \(\lambda_z = 30 - p_z\) users/s.

First, we allow only one arbiter to operate in the market. Interaction with the network is specified as follows. The renting period of network bandwidth (bearer service duration) is one half hour while entry (expansion request) is allowed periodically each 10 minutes. The expansion process is assumed to occur as long as the expected value of profits exceeds its standard deviation. The arbiter aims at maximizing its profit through setting prices to the teleservices and adjusting its bandwidth given that the profit exceeds its standard deviation. We assumed the risk premium to be \(\rho(\phi) = \phi\).

Three hours of network operation was simulated (about two hours CPU time on a SUN workstation). The resulting figures are summarized in Tables 2 and 4. The risk indexes were estimated once every 60 seconds as follows:
\[
\phi_x = \frac{\sum_{t=1}^{60} (r_x - \bar{r}_x)(r_m - \bar{r}_m)}{\sum_{t=1}^{60} (r_m - \bar{r}_m)^2},
\] (14)

where \(r_x\) is the revenue of service class \(z\) at time \(t\). \(\bar{r}_x\) and \(\bar{r}_m\) are the arithmetic averages of \(r_x\) and \(r_m\) respectively and 60 represents the number of samples in our specific example.

The revenue, profit, and surplus are expressed as per-unit-time values. During a time period \(T\), the arbiter's rate of revenue is calculated as
\[
\text{Arbiter's revenue} = \frac{1}{T} \sum_{x \in \{A, B, C, D\}} a_x T_p z_p M_x,
\] (15)

where for class \(x\) users, \(M_x\) and \(p_x\) represent the number of served users and the price charged from them respectively. If \(K\) is the number of channels leased by the arbiter, then the fee paid to the network is \(Kp(K) = 3000K/(750 - 4K)\). Subtracting this fee from the arbiter's revenue, we find the profit which is the arbiter's surplus. The network's and users' surpluses are calculated as follows:
\[
\text{Network's surplus} = \sum_{k=1}^{K} [p(K) - p(k)],
\] (16)

\[
\text{Users' surplus} = \frac{1}{T} \sum_{x \in \{A, B, C, D\}} (\bar{p}_x - p_x) a_x T_p z_p M_x,
\] (17)

where \(\bar{p}_x = (30 + p_x)/2\) is the average reservation price of class \(x\) users.

Next we simulated the network allowing three arbiters to operate in the market. Four hours of network operation was simulated giving the results summarized in Tables 3 and 4.

The example demonstrates that users who are delay-insensitive generate minimum risk for the arbiters, and are charged the least price. Also, big users, which generate less risk, are charged a smaller unit price than

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.00</td>
<td>16.4</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>B</td>
<td>0.31</td>
<td>24.8</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0.16</td>
<td>18.0</td>
<td>0.07</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0.39</td>
<td>22.8</td>
<td>0.22</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3: Results for three-arbiter network.

<table>
<thead>
<tr>
<th>Class</th>
<th>$\phi$ (arb)</th>
<th>$\text{profit} (%)$</th>
<th>Blocking prob.</th>
<th>Waiting time (min)</th>
<th>$\lambda_p$ (users/ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.60</td>
<td>22.7</td>
<td>0</td>
<td>1</td>
<td>18.3</td>
</tr>
<tr>
<td>B</td>
<td>0.49</td>
<td>20.4</td>
<td>0.005</td>
<td>0</td>
<td>9.5</td>
</tr>
<tr>
<td>C</td>
<td>0.14</td>
<td>15.6</td>
<td>0.502</td>
<td>0</td>
<td>14.4</td>
</tr>
<tr>
<td>D</td>
<td>0.27</td>
<td>18.8</td>
<td>0.015</td>
<td>0</td>
<td>11.2</td>
</tr>
</tbody>
</table>

Table 4: Comparison between performance measures.

<table>
<thead>
<tr>
<th>Performance measures</th>
<th>one arbiter</th>
<th>three arbiters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capacity (channels)</td>
<td>76</td>
<td>120</td>
</tr>
<tr>
<td>Utilization (%)</td>
<td>88%</td>
<td>83%</td>
</tr>
<tr>
<td>Revenue</td>
<td>1240</td>
<td>1711</td>
</tr>
<tr>
<td>Profit</td>
<td>728</td>
<td>193</td>
</tr>
<tr>
<td>Users surplus</td>
<td>367</td>
<td>618</td>
</tr>
<tr>
<td>Arborist surplus</td>
<td>738</td>
<td>183</td>
</tr>
<tr>
<td>Network surplus</td>
<td>1259</td>
<td>805</td>
</tr>
<tr>
<td>Total surplus</td>
<td>12114</td>
<td>14058</td>
</tr>
</tbody>
</table>

small users. For example, the unit price of classes $C$ and $D$ are less than that of class $B$. On the other hand, although users of classes $C$ and $D$ have the same size, those of class $C$ are charged less because class $C$ traffic is smoother.

Examining the performance measures, we note that the network and the users are better off under competition; their surpluses are increased. Also, prices and blocking probabilities are significantly improved. Although monopoly results in a higher degree of utilization, it results in a correspondingly higher blocking probability.

In case of monopoly we note that the equilibrium in capacity and prices occur at the point of maximum profit, while under competition, equilibrium occurs at the point of maximum total surplus.

IV. COMMENTS AND CONCLUSIONS

There are basically two time scales in our framework: a macroscopic time scale in which arbiters and network interact, and a microscopic time scale in which users and arbiters interact. An important goal was to prevent connectionless service users, which are characterized by bursty behavior, from stressing the network. This was achieved by leaving this burden to specialized processors (arbiters) which negotiate bandwidth with the network in a relaxed time scale.

Furthermore, it is noteworthy that the proposed scheme adaptively matches the installed capacity to changing traffic conditions. Meanwhile, competition guarantees pushing scarce resources to their best utilization.

The proposed pricing mechanism also distributes network resources among end-users in a competitive way, discouraging users from demanding more than they are actually willing to pay for. Moreover, we have shown that the profits made by arbiters are minimized. This means that network services will be provided to end-users almost transparently and arbiters act only as a tool of managing and dimensioning network resources.

In this paper we addressed a single $o$-$d$ pair. The network case is investigated in [18].

REFERENCES