THRESHOLD PRICING FOR SELLING NETWORK CAPACITY THROUGH FORWARD CONTRACTS

STEVE G. LANNING†, WILLIAM A. MASSEY†, AND QIONG WANG‡

Abstract. In this paper, we consider a telecommunications carrier that serves end users and sells capacities to its peers. We formulate a feedback procedure by which the carrier lets customers bid on capacity contracts and dynamically sets threshold prices to determine which bids to accept. The approach maximizes the carrier’s total revenue from both service and capacity markets. We develop key mathematical techniques for calculating the optimal threshold prices and discuss their properties.

Key words. Bid acceptance, pricing, network capacity, Erlang blocking formula.

AMS(MOS) subject classifications. 91B26, 91B70, 60K26, 90B22, 91B24.

1. Introduction. Network capacity trading among carriers is becoming an industry trend in the telecommunications sector. Efforts have been made to set up exchange markets where carriers sell their capacities as commodities. In other cases, capacity exchange is carried out through forward contracts between interested parties, see Wischik and Greenberg [12] for example.

In this paper, we consider carriers that are only interested in selling capacities. Each carrier decides how to divide its capacity between using some for serving the end users and the rest for exchanging with other carriers. For many carriers, serving end-users is the “bread-and-butter” of their business. These customers arrive to the network at random, acquire a communication channel for a certain period of time, and pay a price based on the duration of the service. We call the revenue from end users service revenue. In the presence of capacity exchange, carriers get additional revenue by selling capacities to others, but doing so reduces their ability to serve end-users, which affects their service revenue. It is in a carrier’s best interest to sell capacities only when the value derived from selling can justify the loss of service revenue.

The value derived from selling network capacities depends strongly on the way the trading is performed. Carriers can auction their capacity to others, in which case the design of auction protocols is important. Carriers can also sell capacities in an exchange market operated by a third party, where they can apply various financial instruments, such as options and futures contracts, to maximize their gains from trading. Furthermore, each

†Aerie Networks, 1400 Glenarm Place, Denver, CO 80202; Email: slanning@aerienetworks.com.
‡Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974; Email: wmassey@lucent.com.
§Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974; Email: chiwang@lucent.com.
carrier can sign a forward contract with another carrier, where the revenue from selling capacity depends on the terms of the contract. All these possibilities cannot be addressed in a single paper, so we limit our study here to the case of a carrier selling its capacity by forward contracting. These contracts specify the amount of network capacity that can be used, as well as the starting and ending times for usage. Interested parties can bid on price. The carrier dynamically sets an unannounced threshold price above which the bid will be accepted. We discuss how to set this threshold price optimally so the carrier can maximize its revenue from both serving end-users and selling capacities to other carriers.

Our problem falls into the general category of revenue management, where a vast amount of research has been performed for the airline industry (for a comprehensive review, see McGill and van Ryzin [10]). The contribution of this paper is to extend the idea of optimizing revenue through capacity allocation to the context of telecommunications systems. (A simple treatment of optimizing profit through capacity allocation for a telecommunication system can be found in Jennings, Massey and McCalla [8] for example.) Specifically, we formulate a contract bidding process that applies to network systems, and develop mathematical machinery to carry out this process.

The paper is organized as follows. Section 2 presents the model formulation and outlines our basic approach. Section 3 describes the key mathematical techniques for solving the problem and formulates the calculation procedure. Numerical examples are demonstrated in Section 4, and concluding remarks are made in Section 5.

2. Formulation and solution approach.

2.1. Problem statement. Consider a carrier who has certain number of communication channels on a single link and operates in two markets:

a. Service market. In the service market, customers start and end the use of a channel at random. The request for usage is either granted instantaneously upon their arrival, if available capacity exists, or blocked otherwise. The customer arrival process is assumed to be Poisson with a constant rate. Each connection lasts for a random period of time and all are assumed to be mutually independent and identically distributed. Moreover, if we assume that the number of channels used by these customers is in steady state, then \( \rho \), the mean offered service load, equals the product of the customer arrival rate and the average duration of the connection. Suppose that the service market is highly competitive, and the carrier is the price taker. The service revenue per unit of time and per channel used is then given exogenously and denoted as \( r \). Let \( \beta \) be the blocking rate, which is determined by the
mean offered service load $\rho$ and the number of available channels. The expected revenue from the service market is then $(1 - \beta) \cdot r \cdot \rho$.

b. **Capacity market.** In the capacity market, customers make advanced reservations for guaranteed use of part of the carrier’s capacity for specific time periods. Suppose that the carrier sells capacity to those customers by signing a contract with them. Each contract allows a customer full usage of a channel from a specific starting time, $T_b$, to a specific ending time, $T_e$. Define $[T_b, T_e]$ as the delivery period and its duration $T_e - T_b$ as the contract length. Note that a customer can reserve multiple channels by signing multiple contracts with the carrier. All contracts are made between time 0 and the starting time of the delivery period $T_b$, where we define the period of $[0, T_b]$ as the contracting window.

During this window, customers arrive randomly according to a Poisson process with mean rate $\lambda_c$. Each arriving customer bids his/her willingness to pay for the contract, and the carrier can either accept or reject the bid. If a bid is accepted, the carrier is obliged to guarantee the customer the use of one channel throughout the delivery period, and the customer is obliged to pay the carrier the accepted bid price. As a convention of forward contracts, the carrier does not need to disclose the contracting information with a customer to a third party (see Chapter 1 of Musiela and Rutkowski [11]). Hence, the carrier’s acceptance or rejection of a bid won’t affect the arrival and bidding strategy of other customers.

Assume the joint distribution of bids is identical, independently distributed with a given probability density function $f(w)$. We also assume that both $\lambda_c$ and $f(w)$ are known to the carrier. If not, the carrier can start with some prior estimates, and use a feedback-based estimation procedure, such as the one described in Keon and Anandlingam [7], to update these parameters.

In general, accepting a bid generates revenue that equals the bid price for the carrier, but reduces revenue from the service market as service customers are then served with less capacity. Therefore, the carrier should accept a bid only if the bid price can at least offset the loss of service revenue. We are interested in developing a policy for the carrier to make this acceptance decision. Specifically, let $L_t (0 \leq t \leq T_b)$ be the number of channels available at time $t$. Suppose a customer shows up at $t$ and bids $w$ for a contract. Should the carrier accept this bid?
2.2. Basic approach. One reasonable approach is to let the carrier set a threshold price \( p(t) \), based on \( L_t \), and accept the bid if and only if the bid price \( w \) exceeds \( p(t) \). This is illustrated in Figure 1 where the \( \times \)'s along the time axis represent the arrival times of all bids. The heights of the vertical line segment above each \( \times \) represents one realization of the random bid price each customer may give. All arriving bids during the time interval \([t, T_b]\) with a price that exceeds the threshold price \( p(t) = p^* \) are represented here by a line segment that touches the shaded region. Such vertical line segments are topped by a black circle whose coordinates are the time of arrival and the bid price. Similar coordinates where the bid price does not exceed the threshold price are marked by an open circle. The number of black circles in the shaded region then correspond to the number of accepted bids.

The value of \( p(t) \) is set to maximize the expected revenue from the capacity market during the remaining contracting period \([t, T_b]\), defined as contracting revenue and denoted as \( E_t[\pi_c] \), and the expected revenue from the service market, defined as service revenue and denoted as \( E_t[\pi_s] \), over the delivery period \([T_b, T_e]\).

Let \( M_t \) be the number of bids that arrive during the remaining contracting period \([t, T_b]\). If \( E[M_t] = \lambda_c \cdot (T_b - t) \) is sufficiently small in comparison with the available capacity \( L_t \), then:

\[
E_t[\pi_c] = \sum_{m=0}^{L_t} Pr(M_t = m) \cdot E[\pi_c(t) | M_t = m]
\]
\begin{equation}
\sum_{m=0}^{L_t} Pr(M_t = m) \cdot m \int_{p(t)}^{\infty} w f(w) dw
\end{equation}

\begin{equation}
\approx E[M_t] \cdot \int_{p(t)}^{\infty} w f(w) dw
\end{equation}

\begin{equation}
= \lambda_c \cdot (T_b - t) \cdot \int_{p(t)}^{\infty} w f(w) dw.
\end{equation}

Given the threshold price \( p(t) \), let \( \Delta L_t \) be the number of bids that are accepted in period \([t, T_b]\). The process that counts the number of contract customer arrivals is Poisson and \( \lambda_c \cdot (T_b - t) \) is sufficiently smaller than \( L_t \), so the distribution of \( \Delta L_t \) is approximately Poisson with a mean of

\begin{equation}
\Lambda_c(t) = \lambda_c \cdot (T_b - t) \cdot \int_{p(t)}^{\infty} f(w) dw.
\end{equation}

Since \( L_b = L_t - \Delta L_t \) equals the number of channels that are available at \( T_b \), and thus are used to serve end users, the revenue from the service market equals:

\begin{equation}
E_s[\pi_s] = \sum_{l=0}^{L_t} Pr(\Delta L_t = l) \left( 1 - \beta(L_b) \right) r \rho \cdot (T_c - T_b)
\end{equation}

\begin{equation}
= \sum_{l=0}^{L_t} \frac{\Lambda_c(t)^l}{l!} e^{-\Lambda_c(t)} \left( 1 - \beta(L_t - l) \right) r \rho \cdot (T_c - T_b)
\end{equation}

\begin{equation}
= \sum_{l=0}^{L_t} \frac{\Lambda_c(t)^l}{l!} e^{-\Lambda_c(t)} B_l,
\end{equation}

where

\begin{equation}
B_l = (1 - \beta(L_t - l)) r \rho \cdot (T_c - T_b)
\end{equation}

and \( \beta(\cdot) \) is the Erlang blocking formula (see Erlang [5]) as a function of the number of channels, with offered load \( \rho \) or

\begin{equation}
\beta(L) = \frac{\rho^L}{L!} \sum_{j=0}^{L} \frac{\rho^j}{j!}.
\end{equation}

To summarize, the bid acceptance decision is now formulated as a problem of finding, for fixed \( t \), a threshold price \( p(t) \) to maximize the following expected revenue function:

\begin{equation}
E_t[\pi] = E_t[\pi_s] + E_t[\pi_c].
\end{equation}
2.3. Optimal threshold price. Below we use the convention that
1/(-1)! = 0. For notational simplicity, we also write \( p(t) \) and \( \Lambda(t) \) as \( p \) and \( \Lambda \) respectively, suppressing their explicit time dependence. Taking derivatives of the expected contracting and service revenue functions with respect to \( p \) gives us:

\[
\frac{dE_t[\tau_c]}{dp} = -\lambda_c \cdot (T_b - t) p f(p) < 0.
\]

Given that \( B_{L_t} = (1 - \beta(0)) \rho \rho (T_c - T_b) = 0 \), we also have

\[
\frac{dE_t[\tau_s]}{dp} = \sum_{i=0}^{L_t} e^{-\Lambda_c} \left[ \frac{N_{i-1} \Gamma_i}{(l - 1)!} - \frac{N_i \Gamma_i}{l!} \right] B_i \frac{d\Lambda_c}{dp}
\]

\[
= -\lambda_c \cdot (T_b - t) f(p) \sum_{i=0}^{L_t-1} e^{-\Lambda_c} \left[ \frac{N_{i-1} \Gamma_i}{(l - 1)!} - \frac{N_i \Gamma_i}{l!} \right] B_i
\]

\[
= \lambda_c \cdot (T_b - t) f(p) \sum_{i=0}^{L_t-1} \frac{N_i \Gamma_i}{l!} (B_i - B_{i+1})
\]

\[
> 0.
\]

From (2.7) and (2.12), increasing \( p \) leads to an increase in the expected service revenue and a decrease in the expected contracting revenue. Therefore, the total expected revenue, \( E_t[\tau_c] + E_t[\tau_s] \), is globally optimized at a unique point \( p^*(t) \) that satisfies:

\[
\frac{dE_t[\tau_c]}{dp} = -\frac{dE_t[\tau_s]}{dp},
\]

i.e.

\[
p^*(t) = e^{-\Lambda_c(t)} r \cdot (T_c - T_b) \rho
\]

\[
\times \sum_{i=0}^{L_t-1} \frac{\Lambda_c(t)^i}{i!} (\beta(L_t - l - 1) - \beta(L_t - l)) \cdot (T_c - T_b)
\]

3. Algorithm. At each time \( t \) within the contracting window \([0, T_b]\), the carrier can determine the optimal threshold price, \( p^*(t) \), from (2.14). Notice that \( \Lambda_c(t) \) depends on \( p^*(t) \), and appears on the right-hand side of the equation. Therefore, an iterative approach such as Newton’s method needs to be applied to solve the equation. The critical step of the approach is to find an efficient technique to evaluate this sum of terms weighted by a Poisson distribution.
Let \( Q \) be a Poisson distribution with mean \( \Lambda_c(t) \) and let \( f(n) = \beta(L - n) \) be the function on the integers that vanishes outside the set \( \{0, 1, \ldots, L\} \). We then have:

\[
(3.1) \quad E[f(Q)] = \sum_{n=0}^{\infty} \frac{e^{-\Lambda_c(t)}\Lambda_c(t)^n}{n!} f(n) = \sum_{n=0}^{L} \frac{e^{-\Lambda_c(t)}\Lambda_c(t)^n}{n!} f(n).
\]

If \( N = \{N(s) \mid s \geq 0\} \) is a Poisson process with rate \( \lambda = \Lambda_c(t) \), then

\[
(3.2) \quad E[f(N(1))] = E[f(Q)].
\]

Notice that \( E[f(N(1) + 1)] - E[f(N(1))] \) is the same as the quantity we want to evaluate. Since \( N \) is a Markov process, we have for all \( f \)

\[
(3.3) \quad \frac{d}{ds} E[f(N(s))] = \lambda \left(E[f(N(s) + 1)] - E[f(N(s))]\right).
\]

We now proceed to give a numerical scheme for computing \( E[f(N(s))] \) and \( E[f(N(s) + 1)] \).

Let \( g_n(s) \equiv E[f(N(s) + n)] \). Using (3.3), we have

\[
(3.4) \quad \frac{d}{ds} g_n(s) = \lambda (g_{n+1}(s) - g_n(s)).
\]

Since \( f \) vanishes on all integers larger than \( L_s - 1 \), we have \( g_{L+1} = 0 \) where \( L_s = L - 1 \). In general, we can compute the vector \( g(s) = [g_0(s), \ldots, g_L(s)] \) by using a forward difference scheme to obtain

\[
(3.5) \quad g_n(s + \Delta s) = (1 - \lambda \Delta s)g_n(s) + \lambda \Delta sg_{n+1}(s).
\]

Note that:

\[
(3.6) \quad g_n(0) = f(n) = \beta(L - n).
\]

Based on (3.5) and (3.6), within a given timestep, we start with \( n = L \), decrease \( n \) down to 0, and obtain the desired result for the case of \( s = 1 \):

\[
(3.7) \quad g_1(1) - g_0(1) = E[f(N(1) + 1)] - E[f(N(1))].
\]

The pseudo-code for this process is as follows:

```plaintext
    time = 0.0;
g(L + 1) = 0.0;
for (n: L; n ≥ 0; n--){
    g(n) = f(n);
}
while (time < 1){
    for (n: L; n ≥ 0; n--){
        g(n) * = (1 - \lambda * \Delta s);
        g(n) += \lambda * \Delta s * g(n + 1);
    }
    time += \Delta s;
}
```

where we must keep \( \Delta s \ll 1/\lambda \).
4. **Numerical examples.** In this section, we demonstrate the application of our approach through numerical examples. We start with a base case scenario in which the offered service load \( \rho = 90 \) and the reward for serving end users per used channel per unit of time is \( \tau = 1 \). Assume that in the capacity market, each contract grants the use of one channel within the period \( [T_b, T_c] \) where \( T_b = 2 \) and \( T_c = 4 \), so the contract length is \( T_c - T_b = 2 \). Within the contracting window \([0, T_b]\), bids for a contract arrive according to a Poisson process with mean rate, \( \lambda_c = 10 \). The bids are exponentially distributed with a mean of 0.2, i.e., \( f(w) = 0.2e^{-0.2w} \).

Let \( L_t \) be the number of channels available at some time \( t < T_b \), where \([t, T_b]\) is the remaining period to make capacity contracts. Figure 2 shows that given the same number of available channels, the larger the value of \( t \) (i.e. a smaller amount of time remains to make contracts), the lower the optimal threshold price, \( p^* \). Nevertheless, \( L_t \) usually decreases as \( t \) increases because more contracts are made during a longer period. The figure shows that \( p^* \) can be higher at a later \( t \) if \( L_t \) is smaller. Therefore, the optimal threshold price may decrease or increase over time.

We now vary some parameters, and discuss the behavior of the optimal threshold price at a given time \( t = 1 \).

4.1. **Effects of service load.** Figure 3 shows the optimal threshold price \( (p^*) \) at \( t = 1 \), given different levels of offered service load, \( \rho \). With higher value of \( \rho \), each channel generates more service revenues, so it is no surprise that \( p^* \) increases with \( \rho \).
It is interesting to see that $p^*$ flattens at both ends of $\rho$. When the network is lightly used by service customers, selling capacity to contract customers does not reduce service revenue significantly. Therefore, the carrier should set $p^*$ to zero and accept all positive bids for capacities. This explains why in Figure 3, there exists a range of low values for $\rho$ where $p^*$ stays close to zero. If the network is heavily loaded with service customers, each channel is used almost all the time. In those situations, selling a channel to contract customers is profitable if and only if the price of the contract exceeds $r(T_e - T_b)$, i.e. the contract length multiplied by the service revenue per channel per unit of time. Hence, $r \cdot (T_e - T_b)$ is the upper limit of $p^*$, and is reached when $\rho$ becomes sufficiently high.

When the network is neither under-loaded nor overloaded, $p^*$ is sensitive to $\rho$. As can be seen from the figure, there exists a range of $\rho$ values, which we define as the critical region, where the optimal threshold price rises rapidly from near zero to its upper-limit. Our approach is the most useful in that region where $p^*$ can be quite different with respect to small changes in $\rho$, and is close to neither zero, nor the upper-limit.

Starting and ending points of the critical region depend on the arrival rate of contracting customers, $\lambda_c$. As shown in Figure 4, with a higher value of $\lambda_c$, the critical region starts and ends at lower levels of $\rho$.

4.2. Optimal threshold price and contract length. We now consider how the optimal threshold price $p^*$ varies with contract length $T_e - T_b$. Intuitively, one might expect that everything else being equal, if the contract length is $k$ times as long as the base case, then a contract costs $k$ times
as much as service revenue. Therefore, the optimal threshold price should be proportional to the contract length. However, our numerical examples show that the conjecture is only true when the network is overly loaded.

Table 1 shows \( p^* \) at \( t = 1 \) given different values of contract length \( T_c - T_b \) and service load \( \rho \). In the base case, we assume \( T_c - T_b = 1 \), and show values of \( p^* \), given \( \rho = 80, 90, 120, 150 \) (assume the number of available channels, \( L_t = 100 \)). We then increase \( T_c - T_b \) from 1 to 8, and show the percentage change of \( p^* \). It can be seen from the table that when \( \rho = 80 \) or 90 (i.e. the service load is 80% or 90% of the available capacity), then the percentage increase of \( p^* \) is slower than the percentage increase of \( T_c - T_b \). When the \( \rho \) reaches 120, 150, \( p^* \) increases linearly with \( T_c - T_b \).

The explanation of this result lies on the fact that increasing \( p^* \) reduces the number of contracts to be accepted, and thus leaves more channels to the service customers. When the network is not over-loaded, increasing the number of channels to accommodate a given service load reduces the channel usage rate. Therefore, the per channel service revenue per unit of time decreases. This means that the cost of selling channels by contract does not increase in proportion to the contract length if the threshold price changes. As a result, \( p^* \) also does not increase proportionally to the contract length. However, if the network is sufficiently overloaded with service customers, the channel usage rate stays at the maximum regardless how many contracts have been accepted. In this case, per channel service revenue per unit of time is insensitive to a change in the threshold price. It

![Figure 4](image-url)
follows that the cost of selling a channel by contract, and thus $p^*$ increases linearly with the contract length.

The above result suggests that in a normally loaded network, a discount in terms of the price per channel per unit of time, should be given to contracts with longer duration.

5. Conclusions. In this paper, we study a mechanism for a communications carrier to sell its network capacity through a forward contract. We develop a threshold pricing approach where the carrier accepts a bid for the contract only when the bidder is willing to pay a price that at least offsets the loss of service revenue due to a smaller number of channels. The approach maximizes the carrier’s total expected revenue from both service and capacity markets. We formulate a feedback-based process for determining the optimal threshold price, and derive key mathematical techniques for price calculation. Under our approach, the optimal price may increase or decrease over time, is close to zero in a lightly loaded network, and reaches its upper-bound in a heavily-loaded network. We also find it is optimal for a carrier to discount contracts with longer duration, unless the network is overloaded.

Our work can be extended in many directions. We can directly apply the analysis of this paper to the case of time inhomogeneous arrivals. We can model $M_t$ as a time inhomogeneous Poisson process so that all increments $M_s - M_t$ for $s < t$ have a Poisson distribution and

\[
E[M_s - M_t] = \int_s^t \lambda_c(\tau) \, d\tau.
\]

Moreover, the resulting $\Delta L_t$ still has a Poisson distribution since it is the “thinning” of the Poisson process $M_t$ (see Daley and Vere Jones [2] for details) where we now have

<table>
<thead>
<tr>
<th>$T_e - T_b$</th>
<th>$\rho = 80$</th>
<th>$\rho = 90$</th>
<th>$\rho = 120$</th>
<th>$\rho = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_e - T_b = 1$</td>
<td>0.310</td>
<td>0.590</td>
<td>0.902</td>
<td>0.984</td>
</tr>
<tr>
<td>2</td>
<td>190%</td>
<td>193%</td>
<td>200%</td>
<td>200%</td>
</tr>
<tr>
<td>3</td>
<td>272%</td>
<td>278%</td>
<td>300%</td>
<td>300%</td>
</tr>
<tr>
<td>4</td>
<td>347%</td>
<td>362%</td>
<td>400%</td>
<td>400%</td>
</tr>
<tr>
<td>5</td>
<td>418%</td>
<td>434%</td>
<td>493%</td>
<td>500%</td>
</tr>
<tr>
<td>6</td>
<td>485%</td>
<td>511%</td>
<td>590%</td>
<td>600%</td>
</tr>
<tr>
<td>7</td>
<td>548%</td>
<td>582%</td>
<td>686%</td>
<td>700%</td>
</tr>
<tr>
<td>8</td>
<td>614%</td>
<td>652%</td>
<td>781%</td>
<td>800%</td>
</tr>
</tbody>
</table>
\( \Lambda_c(t) = \int_t^{T_b} \lambda_c(\tau) \int_{p(\tau)}^{\infty} f(w) dw \, d\tau. \)

Using the modified offered load approximation, see Massey and Whitt [9], we can approximate the blocking rate for the service market by still using the Erlang blocking formula but at time \( t \) we set the new offered load equal to

\[ \rho = E \left[ \int_{t-S}^t \lambda_s(\tau) \, d\tau \right], \]

where \( \lambda_s \) is the mean arrival rate function for the customers in the service market and \( S \) is a random connection holding time. We assume that the holding times of all customers are independent and identically distributed.

Time dependent arrival phenomena, such as a surge in customer traffic at a specific time, can be approximated well by these approximate methods for times of low blocking (under 10%). One effect that arises here is the blocking rate is no longer insensitive to the distribution of the holding time. In Davis, Massey and Whitt [3] we demonstrate this phenomenon for the offered service load. Quantities of interest like the lag between the times of peak demand and the times of peak load will depend on more than the first moments of the holding time. So for a fixed mean holding time, a distribution with a more heavy tailed distribution has a longer lag time. For a discussion of the behavior for the offered service load, see Eick, Massey and Whitt [4].

More general techniques need to be developed to apply our approach to networks with multiple links and multiple-classes of services. It is also interesting to examine a carrier who is not only interested in selling but also buying capacity through forward contracts. Furthermore, we can also look into situations in which a contract can be resold to other bidders, as is the case in the current energy capacity market.

REFERENCES


