# Pricing in a Large Single Link Loss System

Costas A. Courcoubetis[a] and Martin I. Reiman[b]

[a]ICS-FORTH and University of Crete
Heraklion, Crete, Greece
courcou@csi.forth.gr

[b]Bell Labs, Lucent Technologies
Murray Hill, NJ 07974
marty@research.bell-labs.com

We consider a single link loss system where the arrival rates are determined by prices. We examine two economic models, one where prices are set to maximize revenue, and the other where prices are set to maximize total social welfare.

## 1. INTRODUCTION

Loss systems are widely used to study the performance of the provision of guaranteed services by telecommunication networks, where the amount of available resources are finite, and a customer that arrives requesting a certain amount of resources results in either being accepted, or being blocked due to the unavailability of the requested resources. An important performance measure is the probability that such requests are blocked, which depends on the overall request rate, the amount of resources required by the various request types, and the total amount of resources available in the system.

The issue of interest to us here is the role that prices can play in controlling the performance of such blocking systems. A network operator could, by charging customers, influence the rates at which requests for different services arrive (the *demand*), and hence control the way the network will be loaded and the amount of blocking different customers will experience.

In this paper we construct and analyze two economic models that relate to the above observations. For simplicity, both models are discussed in terms of a network consisting of a single link, which has N 'circuits'. The first corresponds to the case of the network manager acting as a monopolist, which sets prices to maximize the total revenue obtained from the system. The second corresponds to the other extreme where the goal of the network manager is to set prices in order to maximize the total utility (*social welfare*) obtained by the users of the network. In this case, the utility of a user is a function capturing his preferences in terms of the frequency of placing calls and the potential negative effects of call blocking. An essential feature of our approach is that it involves an asymptotic analysis, in the regime where both the available capacity and potential load grow to infinity. For the revenue maximization problem we show that, as $N \to \infty$, under

optimal prices the link is either underloaded or critically loaded; it is never overloaded. For the social welfare maximization problem we show that, as $N \to \infty$, the link is always critically loaded. These first order asymptotic results are then supplemented by second order asymptotics for critically loaded systems.

The paper is organized as follows. In section 2 we state the two optimization problems, and in section 3 we review known asymptotic results for the calculation of the blocking probabilities. In sections 4 and 5 we present our results for the revenue and utility maximization problems respectively.

## 2. THE OPTIMIZATION PROBLEMS

The model we consider consists of a link with $N$ circuits and $J$ call types. Calls of type $j$, $1 \le j \le J$, arrive as a Poisson process of rate $g_j$, have mean holding times $\tau_j$, and require $A_j$ circuits (with $A_j \ge 1$ an integer). If there are fewer than $A_j$ idle circuits when a type $j$ call arrives, the call is blocked and lost. We consider two problems that are related to the optimal setting of prices for such a link. The first concerns the maximization of the revenue by the network operator, and the second concerns the optimization of the overall benefit of the customers that use the system (social welfare).

The revenue maximization problem is formulated in terms of demand functions $g_j$, $1 \le j \le J$. Suppose that, given prices $\mathbf{w} = (w_1, \ldots, w_J)$ for the $J$ call types, the arrival rates are $g_1(\mathbf{w}), \ldots, g_J(\mathbf{w})$. Let $B_j(N; \mathbf{g}(\mathbf{w}))$ denote the blocking probability of type $j$ calls in a link with $N$ circuits and arrival rates $g_1(\mathbf{w}), \ldots, g_J(\mathbf{w})$. Each accepted call of type $j$ pays the network operator $w_j$. Blocked calls pay nothing. The revenue maximization problem is the following:

*Maximization of Link Revenue (MLR):*

$$\max_{\mathbf{w}} r(\mathbf{w}) = \sum_{j=1}^{J} w_j g_j(\mathbf{w})[1 - B_j(N; g_1(\mathbf{w}), \ldots, g_J(\mathbf{w}))]\,. \tag{1}$$

Using the known expression for $B_j$, the problem of maximizing link revenue can be turned into a (typically complicated) nonlinear optimization problem. Our goal here is to use asymptotics to gain a better understanding of the structure of the optimal solution. Towards that end we make some intuitively reasonable assumptions about the form of the demand functions.

Let $\rho(\mathbf{w}) = \sum_{j=1}^{J} A_j \tau_j g_j(\mathbf{w})$. For $1 \le J < \infty$, the $J$ dimensional nonnegative orthant is denoted by $\mathbb{R}_+^J = \{\mathbf{x} \in \mathbb{R}^J : 0 \le x_i < \infty,\ 1 \le i \le J\}$, and its interior is denoted by $\mathbb{R}_{++}^J = \{\mathbf{x} \in \mathbb{R}_+^J : 0 < x_i < \infty,\ 1 \le i \le J\}$. We assume that, for each $j$, $1 \le j \le J$, $g_j : \mathbb{R}_{++}^J \to \mathbb{R}_+$ is continuously differentiable, and that, for $\mathbf{w} \in \mathbb{R}_{++}^J$:

(A1) if $g_j(\mathbf{w}) > 0$, then $\dfrac{\partial g_j(\mathbf{w})}{\partial w_j} < 0\,,\ 1 \le j \le J$, (A2) $\dfrac{\partial g_j(\mathbf{w})}{\partial w_i} \ge 0\,,\ 1 \le j \ne i \le J$,

(A3) if $\rho(\mathbf{w}) > 0$, then $\dfrac{\partial \rho(\mathbf{w})}{\partial w_j} < 0\,,\ 1 \le j \le J$, (A4) $\lim\limits_{w \to \infty} \rho(w\mathbf{1}) = 0$,

and

(A5) for any $\epsilon > 0$ there exists a $K_\epsilon < \infty$ such that $\dfrac{\partial g_j(\mathbf{w})}{\partial w_j} > -K_\epsilon$ if $w_j \ge \epsilon_j$, $1 \le j \le J$.

Assumptions (A1) and (A2) are related to monotonicity in demand for individual call types. Assumption (A1) indicates that when the price of a call type increases its arrival rate decreases, while assumption (A2) indicates that there is a possible substitution in the form of an increased arrival rate of other call types. Assumption (A3) states that the total offered load is decreasing in the prices of all call types. Assumption (A4) expresses the natural condition that as all prices increase, the total offered load decreases to zero. Our assumptions do not rule out unbounded demand, but (A5) allows unbounded demand only for very small prices.

The social welfare maximization problem is formulated as follows. There are $J$ call types as before. There are $I$ customers that share the use of the system, and customer $i$, $1 \leq i \leq I$, generates a Poisson stream of calls of type $j$ with rate $g_j^i$, $1 \leq j \leq J$. We let $U^i(g_1^i, \ldots, g_J^i; B_1, \ldots, B_J)$ denote the utility function for customer $i$ (in the simplest case the utility is a function of $g_j^i(1 - B_j)$). The *social welfare* of the above system is defined as the sum of the benefits of the customers, i.e., the quantity $\sum_{i=1}^{I} U^i$.

Since all customers share the link, the blocking probability $B_j$ is the same for all customers, and hence is a function $B_j(N; g_1, \ldots, g_J)$, where $g_j = \sum_{i=1}^{I} g_j^i$ is the total arrival rate of calls of type $j$. The problem of maximizing the social welfare now becomes

*Maximization of Social Welfare (MSW):*

$$\max_{g_j^i, 1 \leq j \leq J, 1 \leq i \leq I} \sum_{i=1}^{I} U^i(g_1^i, \ldots, g_J^i; B_1(N; g_1, \ldots, g_J), \ldots, B_J(N; g_1, \ldots, g_J)) \,. \tag{2}$$

A traditional approach for solving $MSW$ is by using prices. The network operator posts prices $w_j$ for accepting calls of type $j$; the customers adjust their call arrival rates in order to maximize their total net benefit, which is their utility for using the above arrival rates at the particular operating point of the link minus the cost they must pay to the network. An equilibrium in such a system is a set of prices under which the customers do not have the incentive to change their arrival rates $g_j^i$. A desirable property of such an equilibrium is that the corresponding arrival rates also solve $MSW$.

We state some interesting properties that are natural to assume for the utility functions we are using, and which are useful in the proofs of the results in the following sections. It is natural to assume for each customer $i$ that $U^i(x_1, \ldots, x_J; b_1, \ldots, b_J)$ is increasing in the rates $x_j$, and decreasing in the blocking probabilities $b_j$. We also assume that $U^i$ is continuously differentiable in all of its arguments and is concave in each $x_j$. Another property that is natural to assume is that if $x_j(1 - b_j)$ is the 'effective' rate of call arrivals (accepted calls of type $j$) then a customer prefers the same effective rate to occur with less blocking, since blocking can only produce extra overhead. Formally, if $x_j(1 - b_j) = x'_j(1 - b'_j)$ with $b_j \leq b'_j$, then

$$U^i(x_1, \ldots, x_J; b_1, \ldots, b_J) \geq U^i(x'_1, \ldots, x'_J; b'_1, \ldots, b'_J) \,. \tag{3}$$

## 3. CALCULATION OF BLOCKING PROBABILITIES

In this section we provide asymptotic expressions for the blocking probabilities arising in the single link model introduced in the previous section. Although we assume that the system uses the greedy 'complete sharing' admission rule (if there are a sufficient number

of idle circuits at the moment of a call's arrival, then admit the call), our first order asymptotic results hold for more general admission rules as well. It is known that (cf. [3], [7], [4]), under the complete sharing admission rule the stationary distribution of this system is of product form. This typically reduces the calculation of blocking probabilities to the determination of normalizing constants. In this case an efficient one dimensional recursion to calculate $B_j$ is provided by Kaufman [3] and Roberts [7].

### 3.1. First Order Asymptotics

We consider a sequence of systems, indexed by $N$, where $\{A_j, \tau_j, 1 \le j \le J\}$ are held fixed and $g_j(N) \to \infty$ as $N \to \infty$, $1 \le j \le J$. Assume that $g_j(N)/N \to \alpha_j$, $0 < \alpha_j < \infty$, as $N \to \infty$, and let $\rho = \sum_{j=1}^{J} \alpha_j A_j \tau_j$. For $\rho > 1$ let $b(\rho)$ denote the unique root in $[0,1)$ of $\sum_{j=1}^{J} \alpha_j A_j \tau_j (1-b)^{A_j} = 1$. For $0 \le x \le 1$ let $b(x) = 0$. Then, as a special case of a result of Kelly [5] for a network setting we have

$$B_j(N) \to \hat{B}_j(\rho) \equiv 1 - (1 - b(\rho))^{A_j} \quad \text{as} \quad N \to \infty, \quad 1 \le j \le J\,. \tag{4}$$

### 3.2. Second Order Asymptotics

We assume that $\rho = \sum_{j=1}^{J} \alpha_j A_j \tau_j = 1$, and

$$g_j(N) = \alpha_j N + \beta_j \sqrt{N} + o(\sqrt{N}), \quad -\infty < \beta_j < \infty, \quad 1 \le j \le J\,. \tag{5}$$

Using results of Hunt and Kelly [1], it was shown in Reiman [6] that

$$\sqrt{N} B_j(N) \to A_j b^*\,, \tag{6}$$

where $b^* = \sigma^{-1} h(\beta/\sigma)$, $\beta = \sum_{j=1}^{J} \beta_j A_j \tau_j$, $\sigma^2 = \sum_{j=1}^{J} A_j^2 \alpha_j \tau_j$, $h(x) = \frac{\phi(x)}{1-\Phi(x)}$, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2\ 2}$, and $\Phi(x) = \int_{-\infty}^{x} \phi(z)\, dz$.

## 4. MAXIMIZING LINK REVENUE

### 4.1. First Order Asymptotic Analysis

Consider the model of the single link introduced in Section 2, and let the arrival rates be parametrized by $N$, so that $g_j(N, \mathbf{w})$ is the arrival rate of type $j$ calls when the price vector is $\mathbf{w}$, for the link with $N$ circuits. Suppose that

$$N^{-1} g_j(N, \mathbf{w}) \to g_j(\mathbf{w})\,, \tag{7}$$

and let $\rho(\mathbf{w}) = \sum_{j=1}^{J} g_j(\mathbf{w}) \tau_j A_j$. As before let $B_j(N; \mathbf{g}(N))$ denote the blocking probability of type $j$ with $N$ circuits and arrival rates $g_1(N), \ldots, g_J(N)$, where for simplicity we omitted the dependence on the price vector $\mathbf{w}$.

We now derive the asymptotic version of the optimization problem in (1). The revenue of the $N$th system is

$$r_N(\mathbf{w}) = \sum_{j=1}^{J} w_j g_j(N, \mathbf{w}) [1 - B_j(N; g_1(N, \mathbf{w}), \ldots, g_J(N, \mathbf{w}))]\,, \tag{8}$$

and if the normalized revenue is defined as $\tilde{r}_N(\mathbf{w}) = N^{-1} r_N(\mathbf{w})$, by substituting the asymptotic form of the blocking probability from (4), we obtain

$$\lim_{N\to\infty} \tilde{r}_N(\mathbf{w}) = \tilde{r}(\mathbf{w}) = \sum_{j=1}^{J} w_j g_j(\mathbf{w}) \left(1 - b\left(\sum_{j=1}^{J} g_j(\mathbf{w})\tau_j A_j\right)\right)^{A_j} . \tag{9}$$

We next show that the asymptotic form of the $MLR$ problem that results from (9) can be further simplified, since as we show, under the optimal price $\rho \leq 1$ always holds, i.e., the link will never be overloaded.

**Proposition 1** *If* $\mathbf{w}^*$ *maximizes* $\tilde{r}(\mathbf{w})$*, then* $\rho(\mathbf{w}^*) \leq 1$.

**Proof:** To simplify notation we define a normalized version of the quantities $w_j$ and $g_j$, where the corresponding quantities are defined on a per unit of resource usage basis:

$$\hat{g}_j = A_j \tau_j g_j, \qquad \hat{w}_j = \frac{w_j}{A_j \tau_j} . \tag{10}$$

Assume that the optimum occurs for $\rho > 1$ and has the value $R^*$. With no loss of generality let $K$ be the number of distinct prices involved ($K \leq J$), and define by $l_k$ the set of call types priced with $\hat{w}_{l_k}$, where $\hat{w}_{l_1} < \cdots < \hat{w}_{l_K}$. Intuitively, per unit of resource consumption the calls in $l_k$ generate less revenue for the network than calls in $l_{k+1}$.

Now consider the same system where we apply trunk reservation in such a manner that $l_k$ has 'lower priority' than $l_{k+1}$. Let $k^*$ be such that $\sum_{l_k=l_k^*}^{l_K} \sum_{j\in l_k} \hat{g}_j(\hat{\mathbf{w}}) \geq 1$, and $\sum_{l_k=l_{k^*+1}}^{l_K} \sum_{j\in l_k} \hat{g}_j(\hat{\mathbf{w}}) < 1$. It was shown by Hunt and Laws [2] that there are trunk reservation levels such that, with $\rho > 1$, $B_j(N) \to B_j$, $1 \leq j \leq J$, where for $j \in l_k$ with $k < k^*$ $B_j = 1$; for $j \in l_k$ with $k > k^*$ $B_j = 0$; and for $k = k^*$ $0 \leq B_j < 1$. Clearly in this system the revenue is not less than $R^*$ since we admit more expensive calls with higher priority.

We now increase $\hat{w}_{l_1}$ equally for all call types in $l_1$, keeping other prices fixed, until either $\rho$ becomes 1, or $\hat{w}_{l_1} = \hat{w}_{l_2}$. By (A3) $\rho$ decreases as $\hat{w}_{l_1}$ increases. If $\rho = 1$ is reached first, then we have strictly increased the revenue since it must be the case that $k^* = 1$ and (i) the load corresponding to the more expensive types has not decreased because of substitution, (ii) the effective load of $l_1$ being $\sum_{j\in l_1} \hat{g}_j(\hat{\mathbf{w}})(1 - B_{l_1})$ has either remained the same if no substitution occurred (since the new rates $\hat{g}'_j$ at $\hat{w}'_{l_1}$ are such that $\hat{g}'_j = \hat{g}_j(\hat{\mathbf{w}})(1 - B_{l_1})$), or has been substituted by traffic which generates proportionally more reward. If $\hat{w}_{l_1} = \hat{w}_{l_2}$ occurs first, then the revenue is also not decreased, since, because of substitution, the rates of more expensive calls have not been decreased, and since these fill the system, we maintain the rate of revenue we had before. In this case we merge the sets $l_1$ and $l_2$, and repeat the procedure for the $K - 1$ remaining distinctly priced sets. Note that the above argument about increasing the effective arrival rates of more expensive calls is particularly important for the type $l_k^*$ which could be "pushed out" of the system by more expensive types since these might expand because of substitution. In any case the effective part (the proportion that is not blocked) of the rate of $l_k^*$ that generates revenue with $\hat{w}_{l_{k^*}}$ is being replaced by rates of calls that are charged higher prices (per unit of resource usage).

Eventually we will either satisfy the condition $\rho = 1$, or have a price for all call types of $\hat{w}_{l_K}$. If the latter is true, then by increasing the price we will eventually also satisfy $\rho = 1$ by (A4). In any case the corresponding reward will be strictly larger than $R^*$. We must observe that the sequence of intermediate systems are not comparable with the original one since these use trunk reservation, which is one of the causes for producing higher revenue. On the other hand, the final system which operates with $\rho = 1$ is critically loaded, and hence if we switch our policy to complete sharing we will maintain zero blocking probability by (4). Hence the revenue in this last system will be equal to the one obtained with trunk reservation, which is strictly greater than the initial one. This completes our argument. ■

Consider the following non-linear programs

$$\text{Program NMLR}: \text{ maximize } \sum_{j=1}^{J} w_j g_j(\mathbf{w}) \quad \text{subject to } \rho(\mathbf{w}) \leq 1\,, \tag{11}$$

$$\text{Program I}: \text{ maximize } \sum_{j=1}^{J} w_j g_j(\mathbf{w})\,, \tag{12}$$

$$\text{Program II}: \text{ maximize } \sum_{j=1}^{J} w_j g_j(\mathbf{w}) \quad \text{subject to } \rho(\mathbf{w}) = 1\,, \tag{13}$$

and let $(\psi_I^*, \mathbf{w}_I^*)$, $(\psi_{II}^*, \mathbf{w}_{II}^*)$ be the corresponding pairs of the optimal value and argument for the programs $I$ and $II$ respectively. Then if the function $\sum_{j=1}^{J} w_j g_j(\mathbf{w})$ is concave, we obtain the following further simplification.

**Corollary 1.1** *If $\rho(\mathbf{w}_I^*) < 1$ then the solution of NMLR is $(\psi_I^*, \mathbf{w}_I^*)$. If $\rho(\mathbf{w}_I^*) \geq 1$ then the solution of NMLR is $(\psi_{II}^*, \mathbf{w}_{II}^*)$.*

### 4.2. Second Order Asymptotic Analysis

The first order analysis of the previous section yielded (in Proposition 2) $\rho(\mathbf{w}^*) \leq 1$. When $\rho(\mathbf{w}^*) = 1$ it is possible to carry out a more sensitive analysis. Let

$$\mathbf{w}_N = \mathbf{w}^* + N^{-1/2}\mathbf{z} \tag{14}$$

for $\mathbf{z} \in \mathbb{R}^J$. In addition to (7), we now assume that the second order partial derivatives of $g$ are bounded in a neighborhood of $\mathbf{w}^*$. Then Taylor's Theorem yields, for $1 \leq j \leq J$,

$$g_j(\mathbf{w}_N) = g_j(\mathbf{w}^*) + N^{-1/2} \sum_{i=1}^{J} z_i \frac{\partial g_j}{\partial w_i}(\mathbf{w}^*) + o(N^{-1/2})\,. \tag{15}$$

With arrival rates as given in (15), we can use (6) to obtain

$$\sqrt{N} B_j\left(N; g_i(N, \mathbf{w}_N), \ldots, g_J(N, \mathbf{w}_N)\right) \to A_j b^*(\mathbf{z})\,, \tag{16}$$

where $b^*(\mathbf{z}) = \sigma^{-1}h(\beta/\sigma)$, $\sigma^2 = \sum_{j=1}^J A_j^2 g_j(\mathbf{w}^*)\tau_j$, and

$$\beta = \sum_{j=1}^J A_j\tau_j \left( \sum_{i=1}^J z_i \frac{\partial g_j}{\partial w_i}(\mathbf{w}^*) \right) . \tag{17}$$

Let $\hat{r}_N(\mathbf{z}) = N^{-1/2}\left(r_N(\mathbf{w}_N) - r_N(\mathbf{w}^*)\right)$. Combining (8) with (14), (15), and (16), we obtain

$$\hat{r}_N(\mathbf{z}) \to \hat{r}(\mathbf{z}) = \sum_{j=1}^J \left[ z_j g_j(\mathbf{w}^*) + w_j^* \sum_{i=1}^J z_i \frac{\partial g_j}{\partial w_i}(\mathbf{w}^*) - w_j^* g_j(\mathbf{w}^*) A_j b^*(\mathbf{z}) \right] . \tag{18}$$

The entire second order analysis is predicated on the assumption that $\rho(\mathbf{w}^*) = 1$. Thus $\mathbf{w}^*$ solves Program II, given by (13). Solving Program II with a Lagrangian analysis, allows us (after some manipulation) to conclude that

$$\hat{r}(\mathbf{z}) = \mu \sum_{i=1}^J \sum_{j=1}^J z_i A_j \tau_j \frac{\partial g_j(\mathbf{w}^*)}{\partial w_i} - \sum_{j=1}^J w_j^* g_j(\mathbf{w}^*) A_j b^*(\mathbf{z}) .$$

Note that we can write $\hat{r}(\mathbf{z}) = f(\beta(\mathbf{z})) \equiv \mu\beta(\mathbf{z}) - \hat{b}(\beta(\mathbf{z})) \sum_{j=1}^J w_j^* g_j(\mathbf{w}^*) A_j$, where $\beta$ is given in (17) and $\hat{b}(\beta) = \sigma^{-1}h(\beta/\sigma)$. Thus, the second order optimization problem, which entails the maximization of $\hat{r}(\mathbf{z})$ over $\mathbf{z} \in \mathbb{R}^J$, can be reduced to the problem of maximizing $f(\beta)$ over $\beta \in \mathbb{R}$. Given a $\beta^*$ such that $f(\beta^*) \geq f(\beta)$ for any $\beta \in \mathbb{R}$, we are free to choose any $\mathbf{z}^*$ for which $\beta(\mathbf{z}^*) = \beta^*$.

By the Lagrangian analysis, $\mu \geq 0$. In addition, $\hat{b}$ is strictly convex and strictly increasing. When $\mu > 0$ there is a unique (finite) $\beta^*$ that maximizes $f(\beta)$. When $\mu = 0$ this analysis yields $\beta^* = -\infty$, which implies that an even more sensitive analysis is needed to deal with this case. We do not pursue this issue further in this paper.

## 5. MAXIMIZING SOCIAL WELFARE

The problem of the social planner is to choose the arrival rates $g_j^i$, $1 \leq i \leq I$, $1 \leq j \leq J$, in order to maximize the sum of the utilities of all the customers (social welfare). An implicit way to do this is through prices: the social planner posts prices for each of the call types, under which the customers, by doing their local optimization, will choose the arrival rates that correspond to the optimal solution of the social welfare problem. The first question we answer is if such a set of prices exists.

Consider the social welfare function $W = \sum_{i=1}^I U^i(g_1^i, \ldots, g_J^i; B_1, \ldots, B_J)$, where $B_j = B_j(g_1, \ldots, g_J)$ is the blocking probability of calls of type $j$, and $g_j = \sum_{i=1}^I g_j^i$ is total arrival rate of calls of type $j$. Then at the optimum

$$\frac{\partial W}{\partial g_j^i} = \frac{\partial U^i}{\partial g_j^i} + \sum_{k=1}^I \sum_{l=1}^J \frac{\partial U^k}{\partial B_l} \frac{\partial B_l}{\partial g_j^i} = 0 \, , 1 \leq i \leq I, \; 1 \leq j \leq J,$$

and since $B_l$ depends on $g_j^i$ through the sum $g_j = \sum_{i=1}^I g_j^i$, it follows that the above condition becomes

$$\frac{\partial U^i}{\partial g_j^i} + \sum_{k=1}^I \sum_{l=1}^J \frac{\partial U^k}{\partial B_l} \frac{\partial B_l}{\partial g_j} = 0 \, , \; 1 \leq i \leq I, \; 1 \leq j \leq J. \tag{19}$$

Let $w_1, \ldots, w_J$ be the prices charged to the customers for the accepted calls (which are not blocked). Then customer $i$ will choose arrival rates that solve the local optimization problem

$$\max_{g_j^i, j \in J} U^i(g_1^i, \ldots, g_J^i; B_1, \ldots, B_J) - \sum_{k=1}^{J} w_k g_k^i (1 - B_k)\,, \tag{20}$$

where the values of the blocking probabilities $B_k$ are those corresponding to the current operating point of the link and are considered as given (measured). This is an important assumption which leads to the above definition of the local optimization problem. The case in which the users have knowledge of the derivatives of the blocking probabilities with respect to their arrival rates leads to a different optimization problem, in which such prices might not exist. In any case, if the size of the system is large compared to individual users, then a reasonable approximation is that an individual user cannot have a significant effect on the blocking that takes place, and hence such a user is faced with solving (20).

The arrival rates chosen will satisfy the conditions

$$\frac{\partial U^i}{\partial g_j^i} - w_j(1 - B_j) = 0,\ 1 \le j \le J, \quad 1 \le i \le I\,. \tag{21}$$

Observe now that if we choose the prices $w_j^*$ so that

$$w_j^* = -(1 - B_j)^{-1} \sum_{k=1}^{I} \sum_{l=1}^{J} \frac{\partial U^k}{\partial B_l} \frac{\partial B_l}{\partial g_j}, \quad 1 \le j \le J, \tag{22}$$

we are guaranteed that the global conditions (19) and the local conditions (21) are equivalent, and hence the optimal allocation of arrival rates for $MSW$ is an equilibrium for the system of prices (22). Note that the above prices are "congestion" prices, in the sense that a customer pays the rest of the customers (including himself) for the marginal decrease of their utility due to the increase of blocking that is produced because of his increased rate of requests.

The form of (22) suggests that in order to compute the above prices we need the explicit knowledge of the utility functions of the customers. An interesting observation is that if the customers of the network fall into a small number of generic classes for which the utilities can be assumed known, then the network operator only needs to know the number of users of each particular class to compute the above prices.

### 5.1. First Order Asymptotic Analysis

We consider the following asymptotic regime. There are $I$ classes of customers and $J$ types of calls, with these quantities held fixed. In the system with $N$ circuits we have $N_i$ customers of class $i \in I$, where $N_i = \lfloor N\theta_i \rfloor$, $\sum_{i=1}^{I} \theta_i = \theta$, $0 < \theta < \infty$. A customer of class $i \in I$ has a utility function $U^i$ as in section 2, which is strictly increasing and concave in $g_j^i$, $j \in J$. Then the problem of maximizing the social welfare function is

$$\max_{g_j^{ik}, i \in I, k \in N_i, j \in J} W = \sum_{i=1}^{I} \sum_{k=1}^{N_i} U^i(g_1^{ik}, \ldots, g_J^{ik}; B_1, \ldots, B_J)\,, \tag{23}$$

and since the utilities are concave in $g_j^{ik}$, at any optimum we must have $g_j^{ik} = g_j^{ik'}$, $1 \le k, k' \le N_i$, $1 \le i \le I$. Hence if we denote the above arrival rates by $g_j^i$, (23) is equivalent to

$$\max_{g_j^i, i\in I, j\in J} W = \sum_{i=1}^{I} N_i U^i(g_1^i, \ldots, g_J^i; B_1, \ldots, B_J)\,.$$

Now, after normalizing by dividing by $N$ and taking the limit as $N \to \infty$, we obtain that in the limit the optimization problem for $\tilde{W} = \lim_{N\to\infty} N^{-1}W$ becomes

$$\max_{g_j^i, i\in I, j\in J} \tilde{W} = \sum_{i=1}^{I} \theta_i U^i(g_1^i, \ldots, g_J^i; \hat{B}_1(\rho), \ldots, \hat{B}_J(\rho))\,, \tag{24}$$

where $\hat{B}_j(\rho)$ is given by (4), using $\alpha_j = g_j = \sum_{i=1}^{I} \theta_i g_j^i$.

If property (3) holds, the following proposition states that (24) can be further simplified, since it never pays to have blocking.

**Proposition 2** *If $y_j^i$, $i \in I$, $j \in J$ maximize $\tilde{W}$ in (24), and (3) holds, then $\rho = 1$.*

**Proof:** Assume first that $y_j^i$, $i \in I$, $j \in J$ are such that $\rho < 1$. Then the blocking probabilities from (4) are all zero, and $\tilde{W} = \sum_{i=1}^{I} \theta_i U^i(y_1^i, \ldots, y_J^i; 0, \ldots, 0)$. Now since the utilities are increasing functions of the arrival rates, we can always increase some rate, say $y_j^i$, while keeping $\rho < 1$, and since the blocking probabilities will stay zero, we will get a strict increase of $\tilde{W}$, which shows that optimality can not be achieved when $\rho < 1$.

Suppose now that $y_j^i$, $i \in I$, $j \in J$ are such that $\rho > 1$. Then the blocking probabilities from (4) are all positive. Consider the point $y_j^i(1 - B_j)$, $i \in I$, $j \in J$. Clearly at this point $\rho = 1$, and hence the blocking probabilities are zero. Hence by using (3) we get that at this new point the social welfare is not decreased from $\tilde{W}$. This completes the proof. ■

**Corollary 2.1** *The maximization of the social welfare corresponds to the program*

$$\textit{maximize} \quad \sum_{i=1}^{I} \theta_i U^i(g_1^i, \ldots, g_J^i; 0, \ldots, 0) \tag{25}$$

$$\textit{subject to} \quad \sum_{j=1}^{J} \sum_{i=1}^{I} \theta_i g_j^i \tau_j A_j \;=\; 1. \tag{26}$$

Solving (26) is now straightforward. In order to obtain a better insight, we use the normalized version of $g_j^i$ as defined in (10). Then (26) becomes

$$\text{maximize} \quad \sum_{i=1}^{I} \theta_i \hat{U}^i(\hat{g}_1^i, \ldots, \hat{g}_J^i; 0, \ldots, 0) \tag{27}$$

$$\text{subject to} \quad \sum_{j=1}^{J} \sum_{i=1}^{I} \theta_i \hat{g}_j^i = 1. \tag{28}$$

A simple application of Lagrangian methods indicates that at the optimum $\frac{\partial \hat{U}^i}{\partial \hat{g}_j^i} = w$ for all $i \in I$, $j \in J$, and that this is achieved with $w$ being the price per request of all types. Because of the normalization we have used, this implies that the optimal price is such that requests are charged proportionally to the amount of resources they consume. (This will typically not be the case in the revenue maximization problem unless the demand functions have a special structure.)

### 5.2. Second Order Asymptotic Analysis

As in the revenue maximization problem, it is possible to carry out a more sensitive analysis for the social welfare maximization as well. As seen in Proposition 3, $\rho = 1$ always holds in the first order asymptotic analysis for social welfare maximization.

Let

$$g_j^i(N) = g_j^i + N^{-1/2} z_j^i, \quad 1 \le i \le I,\ 1 \le j \le J\,, \tag{29}$$

where $\{g_j^i,\ 1 \le i \le I,\ 1 \le j \le J\}$ is a solution of (25) and (26). By (6) we obtain

$$\sqrt{N} B_j(N_j g_i(N),\ \ldots,\ g_J(N)) \to A_j b^*(\mathbf{z})\,, \tag{30}$$

where $\mathbf{z} = (z_j^i,\ 1 \le i \le I,\ 1 \le j \le J) \in \mathbb{R}^{I \times J}$, $b^*(\mathbf{z}) = \sigma^{-1} h(\beta(\mathbf{z})/\sigma)$, $\sigma^2 = \Sigma A_j^2 g_j \tau_j$, and $\beta(\mathbf{z}) = \sum_{j=1}^J A_j \tau_j \sum_{i=1}^I \theta_i z_j^i$. We assume that the second order partial derivatives of $U^i$ are bounded in a neighborhood of $(g_1^i,\ \ldots,\ g_J^i;\ 0,\ \ldots,\ 0)$. Then again applying Taylor's Theorem and defining $\hat{W}_N(\mathbf{z}) = N^{-1/2}\,(W(\mathbf{g}(N)) - W(\mathbf{g}))$ yields

$$\hat{W}_N(\mathbf{z}) \to \hat{W}(\mathbf{z}) = \sum_{i=1}^I \theta_i \sum_{j=1}^J \frac{\partial U^i}{\partial g_j^i} z_j^i + b^*(\mathbf{z}) \sum_{i=1}^I \theta_i \sum_{j=1}^J A_j \frac{\partial U^i}{\partial B_j}\,. \tag{31}$$

From the first order asymptotic analysis, $\frac{\partial U^i}{\partial g_j^i} = w \tau_j A_j,\ \ 1 \le i \le I,\ 1 \le j \le J$ for some $0 < w < \infty$. Thus we can write

$$\hat{W}(\mathbf{z}) = w\beta(\mathbf{z}) + \hat{b}\,(\beta(\mathbf{z})) \sum_{i=1}^I \sum_{j=1}^J \theta_i A_j \frac{\partial U^i}{\partial B_j}\,,$$

reducing the second order optimization problem to a one dimensional unconstrained problem as in Section 4.2.

## REFERENCES


1. P. J. Hunt and F. P. Kelly. On critically loaded loss networks. *Adv. Appl. Prob.*, 21:831–841, 1989.
2. P. J. Hunt and C. N. Laws. Optimization via trunk reservation in single resource loss systems under heavy traffic. *Ann. Appl. Prob.*, 7:1058–1079, 1997.
3. J. S. Kaufman. Blocking in a shared resource environment. *IEEE Trans. Comm.*, COM-29:1474–1481, 1981.
4. F. P. Kelly. *Reversibility and Stochastic Networks.* Wiley, New York, 1979.
5. F. P. Kelly. Blocking probabilities in large circuit switched networks. *Adv. Appl. Prob.*, 18:473–505, 1986.
6. M. I. Reiman. A critically loaded multiclass Erlang loss system. *Queueing Syst.*, 9:65–82, 1991.
7. J. W. Roberts. A service system with heterogeneous user requirements – application to multi-services telecommunications systems. In *Performance of Data Communication Systems and Their Applications.* North-Holland, 1981).