Analysis And Provisioning of a Circuit-switched Link with Variable-Demand Customers

Wenhong Tian and Harry Perros

Department of Computer Science, North Carolina State University, NC 27606, USA

Abstract. We consider a single circuit-switched communication link, depicted by a Erlang multi-class loss queue, where a customer may vary its required bandwidth during its service. We obtain approximately the steady-state blocking probability of each class of customer. Comparisons with simulation results show that the approximation solution has a good accuracy. For the proposed model, we also provide an efficient capacity provisioning algorithm.

1 Introduction

In circuit-switched communication systems and connection-oriented packet switched networks, a connection is typically allocated a fixed bandwidth which does not vary over the life of the connection. However, in today's dynamically changing communication networks, the bandwidth allocated to a connection may have to vary in order to accommodate load fluctuations. In this paper, we consider the case where the customer's bandwidth requirements change during its service time. This case has been motivated by the Link Capacity Adjustment Scheme (LCAS) in the Data over SONET (Dos) architecture.

Traditional SONET/SDH was optimized to carry voice traffic. It was also defined to carry ATM traffic and IP packets (PoS). Changes in the capacity allocated to a connection are done manually. Recently, a novel architecture has been proposed, referred to as data over SONET/SDH (DoS) which provides a mechanism for the efficient transport of integrated data services.

It utilizes three schemes, namely, the Generic Frame Procedure (GFP), Virtual Concatenation (VCAT), and Link Capacity Adjustment Scheme (LCAS) [13]. GFP is a simple adaptation scheme that extends the ability of SONET/SDH to carrying different types of traffic. Specifically, it permits the transport of frame-oriented traffic, such as Ethernet and IP over PPP. It also permits continuous-bit-rate block-coded data from Storage Area Networks (SAN) transported by networks, such as Fiber Channel, Fiber Connection (FICON), and Enterprise System Connect (ECON).

Virtual concatenation maps an incoming traffic stream into a number of individual subrate payloads. The subrate payloads are switched through the SONET/SDH network independently of each other (see for example, Perros [8]).

Virtual concatenation is only required to be implemented at the originating node where the incoming traffic is demultiplexed into subrate payloads and at the terminating node, where the payloads are multiplexed back to the original stream. The individual payloads might not necessarily be contiguous within the same OC-N payload. Finally, the number of subrate payloads allocated to an application is typically determined in advance. However, the transmission rate of the application may vary over time. In view of this, it can be useful to dynamically vary the number of subrate payloads allocated to an application. This can be done using the link capacity adjustment scheme (LCAS). In LCAS, signaling messages are exchanged between the originating and terminating SONET/SDH node to determine and adjust the number of required subrate payloads. LCAS makes sure that the adjustment process is done without losing any data.

The calculation of call blocking probabilities in circuit-switched networks has been extensively analyzed. However, this has been done under the assumption that the bandwidth allocated to a customer does not change throughout the customer's service. For instance, Kaufman [5], Roberts [9], and Nilsson et al.[7] developed efficient algorithms for calculating the blocking probabilities of a multi-rate loss queue. In this case, customers belong to different classes and each class is associated with a class-dependent arrival rate, class dependent service rate, and a class-dependent bandwidth requirement expressed in number of servers. However, a class r customer cannot switch classes during its service time, and as a result, it cannot change the number of servers allocated to it. Call blocking probabilities over an entire circuit-switched network have been computed under a variety of assumptions, see for instance Kelly [6], Ross [10], Alnowibet and Perros [1], Washington and Perros [12], under assumptions similar to the above case of a single loss queue.

In this paper, we consider the multirate single loss queue depicting a circuit-switched communication link, this link may be an optical link, or a wired or wirelss TDM link. Each server represents a time slot in a TDM link or a subrate stream in an optical link. Class r calls arrive in a Poisson fashion at the rate of λ_r, and require initially b_r servers. During the service time of the call, the number of servers required may change. The call is not blocked if fewer servers than currently allocated to it are required. However, the call will get blocked if additional servers are required and these servers are not available at that instance. We describe an approximation algorithm for the calculation of the call blocking probability of each class. To the best of our knowledge this queueing system has not been analyzed before.

We also use a provisioning method based on Hampshire et al. [4] to determine the minimum number of required servers.

This paper is organized as follows. In section 2, we describe in detail the multi-class loss queue under study and how it can be used to model various cases where a customer may change its bandwidth requirements during its service. In section 3, we describe the approximation algorithm and in section 4, we describe how to calculate the minimum number of servers of the loss system so that the blocking probability of any class is less than a pre-specified value. Numerical examples are given in section 5, and finally the conclusions are given in section 6.

2 The Multi-Class Loss Queue with Variable-Demand Customers

Let us consider a multi-class loss queue. There are R classes of traffic, and class i customers (i=1,2,...,R) arrive at the loss queue in a Poisson fashion with a class-dependent arrival rate λ_i requiring b_i servers. Class i customers receive an exponentially distributed service time with mean $1/\mu_i$. The required number of servers is ordered for convenience, that is, $0 < b_1 \leq b_2 \ldots \leq b_R$. Upon arrival at the loss queue, a customer is blocked if the required number of servers is not available. After an exponentially distributed service with rate μ_i, a class i customer may depart from the system with probability p_{i0} or it may change its class to class k with probability p_{ik} where $\sum p_{ij} = 1$. A class change implies that the customer's bandwidth requirements change from b_i to b_k. If $b_k < b_i$, then the class change is successful and the $b_i - b_k$ remaining unused servers join in the pool of available servers. However, if $b_k > b_i$, then $b_k - b_i$ additional servers are required. The customer is blocked (i.e. lost) if these $b_k - b_i$ additional servers are not all available at that moment.

Define P1(i,j)=p_{ij} which is a matrix with size $R \times (R+1)$. Let P be submatrix of P1 and P has dimension $R \times R$. We have to assure that $(I - P)$ is invertible so that a customer entering the system eventually exists.

As will be seen below, the analysis of this system permits a large number of servers which allows us to model high-bandwidth circuit-switched links. For instance, an OC-768 link will be modeled by a loss queue with 768 servers where a server represents an OC-1 subrate stream. The analysis of this model also permits a large number of classes. This feature gives the model the required flexibility to depict users with a given bandwidth profile. For instance, let us consider an example where one group of users requires initially 10 servers. This bandwidth requirement is changed to 20 servers and after that to 15 servers. Then, we say that the bandwidth profile of this group of users is: $\{10, 20, 15\}$. This will be modeled using three classes, say 1, 2, and 3, as follows. A customers with this profile arrives at the loss queue with an arrival rate λ_1 (arrival rates of λ_2 and λ_3 are equal to zero) requiring b_1 = 10 servers. After an exponential service time with a mean of $1/\mu_1$, a customer changes to class 2, thus requiring b_2 = 20 servers with probability p_{12}=1. Following an exponential service time with a mean of of $1/\mu_2$, the customer changes to class 3 with probability p_{23}=1. Finally, after an exponential service time with a mean of $1/\mu_3$, the customer departs, i.e. p_{30}=1. A customer changing from class 1 to 2, may get blocked if the additional 5 servers are not available. However, a customer going from class 2 to 3 will never get blocked since it requires fewer servers than those it held. More complex bandwidth profiles can be constructed by selecting a set of unused classes, and associating each class i with a set of values for $1/\mu_i$, b_i, p_{ij}. Each set of classes associated with a specific bandwidth profile can be seen as forming a closed super-class within which class changes are allowed in a pre-specified manner. The case where a customer can change bandwidth requirements in a random manner can be readily accommodated.

3 Calculation of call blocking probabilities

For the classical multi-class loss system without bandwidth adjustments, there are well-known results.

Let us assume that the system has a total of C identical servers (channels or units of bandwidths), and each can provide service to any class of arrivals. Let n=$(n_1, n_2, ..., n_R)$ where n_r is the number of class r customers in the system, and let b=$(b_1, b_2, ..., b_R)$.

The total number of busy servers in state n is

$$bn^T = b_1 n_1 + b_2 n_2 + ... + b_R n_R. \tag{1}$$

The set of all possible states of the system can be described as

$$S^b = \{n : bn^T \leq C\}. \tag{2}$$

It is well known that the multi-class loss system has a product-form solution given by:

$$P(n) = \prod_{i=1}^{R} \frac{\rho_i^{n_i}}{n_i!} G^{-1}(\Omega), \forall n \in \Omega \tag{3}$$

where

$$G(\Omega) = \sum_{n \in \Omega} \prod_{i=1}^{R} \frac{\rho_i^{n_i}}{n_i!} \tag{4}$$

and $\rho_i = \lambda_i / \mu_i$. The challenge is to obtain the blocking probability for each class. Computing the blocking probabilities by directly enumerating all possible states of the system requires an $O(C^R)$ amount of time. The direct method is computationally cumbersome and grows exponentially fast even for relatively small systems. Several methods have been presented in the literature to avoid the exponential complexity of the computations. One of the most powerful methods for obtaining the blocking probabilities was published independently by Kaufman (1981) [5] and Roberts (1981) [9]. The Kaufman-Roberts method is a fast recursive algorithm that has a linear complexity of $O(CR)$. The recursive formula is as follows:

$$w(k) = \frac{1}{k} \sum_{r=1}^{R} \rho_r b_r w(k - b_r), k = 1, 2, ..., C. \tag{5}$$

where w(0)=1 and $\rho_r = \lambda_r / \mu_r$. Then, the blocking probability of class r arrivals is given by:

$$B_r = \frac{\sum_{j=C-(b_r-1)}^{C} w(j)}{\sum_{j=0}^{C} w(j)}, r = 1, 2, ..., R. \tag{6}$$

It is interesting to know that this formula can be applied to the single class model, as a fast way of obtaining the blocking probability. Given the blocking probabilities, the average number of class r customers in the system is

$$E[Q_r] = \rho_r (1 - B_r), r = 1, 2, ..., R. \tag{7}$$

The multi-rate loss model with variable-demand customers described in the previous section can be analyzed numerically by setting up the underlying rate matrix and subsequently solving it in order to obtain the stationary probability vector. However, this numerical approach is limited to small size problems due to the complexity involved in setting up the rate matrix. It is also difficult to obtain a closed-form expression because of the variable-demand customers.

In view of these considerations, we solve this loss system approximately as follows.

We assume that when a customer changes its class from i to another class, say class j, it simply departs from the loss queue and it re-joins it as a new class j customer. Its departure from the loss queue and its arrival to the loss queue are not synchronized. That is, we simply calculate a new arrival rate for class i customers based on the external arrival rate and all the possible feedbacks due to customers changing their class to class i. Specifically, we have that the departure rate of class i customers from the loss model is:

$$\mu_i E[Q_i] = \mu_i \rho_i (1 - B_i) = \lambda_i (1 - B_i) \tag{8}$$

Then, the total class i arrival rate due to feedback from other classes is:

$$\lambda_{hi} = \sum_{k=1}^{R} \bar{\lambda_k} (1 - B_k) p_{ki} \tag{9}$$

where p_{ki} is the probability that a class k customer will change to class i and $\bar{\lambda_k}$ is the total class k effective arrival rate (i.e., external arrival rate plus feedbacks from the other classes).

Thus the total effective arrival rate $\bar{\lambda_i}$ of class-i to the loss model is:

$$\bar{\lambda_i} = \lambda_i + \lambda_{hi} = \lambda_i + \sum_{k=1}^{R} \bar{\lambda_k} (1 - B_k) p_{ki} \tag{10}$$

where λ_i is the class i external arrival rate to the loss queue.

This equation is often called the traffic equation. The effective arrival rate and the blocking probability of each class are unknown and have to be decided iteratively.

The total offered load for each class is given by the following nonlinear matrix equations obtained from (10):

$$\bar{\rho} = (I - P^T \bar{B})^{-1} \rho \tag{11}$$

where I is the identity matrix, $\bar{B}$=diag($[1 - B_1, 1 - B_2, \ldots, 1 - B_R]$) is a diagonal matrix, P is the class-changing probability matrix with its elements $P(i, j)$=p_{ij}, $\bar{\rho}$=$[\bar{\rho_1}, \bar{\rho_2}, ..., \bar{\rho_R}]$ where $\bar{\rho_i}$=$\bar{\lambda_i}/\mu_i$ is the effective offered load of class i and ρ=[ρ_1 , ρ_2,..., ρ_R]. We can now use (11) in expression (5) in order to calculate the class blocking probabilities.

The resulting system of equations is solved by a fixed-point procedure summarized below.

Summary of algorithm

set small value for degree of accuracy ϵ

do (the following steps)

Step 1: Set initially all blocking probabilities and λ_{hi} to be zero

Step 2: compute values for total offered load per ρ_i=λ_i/μ_i

Step 3: compute values for blocking probabilities B_i per (6)

Step 4: update values for total offered load per (11)

Step 5: update values for blocking probabilities per (6)

while (relative error of two successive blocking prob.$>\epsilon$)
end while

The above algorithm for the calculation of call blocking probabilities has a time complexity of $O(log2(CR/\epsilon)CR^2)$. This algorithm is scalable in the number of classes. A proof of convergence and complexity for a single-class traffic using the bisection algorithm has been sketched out in [11]. For multi-class case considered in this paper, we do not provide a proof of convergence. Through numerous examples, however, we observed that this algorithm converges very fast, often in a few tens of steps.

As is known, the Kaufman-Roberts algorithm is numerically unstable, i.e., it causes overflows when the offered load and/or the total number of servers is very large. This can be avoided by using a dynamic factoring technique. We use a small number α as a scaling factor to avoid potential overflows. If upon inspection, it is found that an overflow would occur in the computation of $w(k)$ in the Kaufman-Roberts formula, all $w(i)$ are scaled, i.e., w(i)=w(i)α for i=$0, 1, ..k$, so that each w(i) is small enough. The process of dynamic scaling increases the computational costs, but the order of the overall complexity remains unchanged.

We observed that a simpler algorithm can be used in the following two cases:

1). When the total number of servers is very large comparing to the offered loads so that the blocking probability for each class is very small (for example, less than 0.001), equation (11) can be approximated by $\bar{\rho} = (I - P^T)^{-1}\rho$, i.e., we can set all the blocking probabilities equal to zero. In this case, we can calculate the effective offered load from (11) without iterating on the blocking probabilities.

2). If total capacity C is very large, then the feedback rate from class i to any class j can be simply expressed as $\lambda_i p_{ij}$. In this case, the solution is simplified as in case (1) above.

The above two cases provide an upper bound on the effective offered load.

4 Capacity Provisioning

Provisioning optimal total capacity is one of practical ways to meet the blocking probability and other QoS requirements. In this section, we describe how to calculate the minimum number of servers C of the loss model so that the maximum blocking probability of any class is less than a pre-specified value ϵ for a given load. This permits the blocking probabilities of the remaining classes to also be less than ϵ.

This minimum value of C can be calculated iteratively using the fixed-point algorithm described in the previous section. However, when the required capacity C is very large, this iterative approach becomes CPU intensive since its time complexity is $O(log2(CR/\epsilon)CR^2)$.

It is a long-standing conjecture that the optimal number of servers is of the form ρ+K$\sqrt{\rho}$ for single class traffic where K is a constant depending on the offered load and blocking probability. This approximation yields very accurate results. Indeed, based on extensive sensitivity tests, the actual optimum and approximate values rarely deviate by more than one server, or by more than one percent, whichever is greater (see Grassman

[3]). Hampshire et al. [4] obtained the following asymptotic expression for the optimum value of C in the multiclass case:

$$C = \sum_{i=1}^{R} b_i \bar{\rho}_i + \psi\left(\min_{1 \le i \le R} \frac{\epsilon_i}{b_i} \sqrt{\sum_{i=1}^{R} b_i^2 \bar{\rho}_i}\right) \sqrt{\sum_{i=1}^{R} b_i^2 \bar{\rho}_i} \quad (12)$$

where ϵ_i is the blocking probability requirement for class i and $\psi(x)$ is the unique solution of the following differential equation

$$\psi'(x) = \frac{-1}{(\psi(x) + x)x}, \psi(\sqrt{2/\pi}) = 0 \quad (13)$$

In [4], the authors suggest to use a lookup table for values of $\psi(x)$ by using a second order Runge-Kutta method to compute $\psi(x)$. However, a lookup table may not be practical if the step size is very small and we do not know the starting point x. In this paper, we have solved equation (13) to obtain

$$x^{-1} e^{-0.5\psi(x)^2} - \sqrt{2\pi} erf(0.5^{1/2}\psi(x)) - x\sqrt{0.5\pi} = 0 \quad (14)$$

where $erf(.)$ function is defined as follow:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x exp(-t^2) dt \quad (15)$$

Given x, equation (14) can be easily solved numerically for $\psi(x)$. Applied to equation (12), we obtain the requested total capacity. Because of the asymptotic rule [4], satisfying the requirements provides more than enough capacity for all the other classes. Through many numerical examples, we observed that the minimum capacity C obtained using equation (12) is very closed to the exact solution.

5 Numerical examples

In this section, we validate the accuracy of our approximation and provide some insights into the multi-rate loss queue with variable-demand customers. We also provide a capacity provisioning example.

The approximation results were compared against simulation data. 95% confidence intervals were also calculated, but since they are extremely small, they are not given in the results below. In Table 1, we give the approximate and simulation results of call blocking probabilities for three classes customers with the number of servers C varying from 20 to 50. The following parameters were used: ρ=[1,2,3],b=[1,2,3], p_{i0}=0.5, i=1, 2, 3. Any class i customer can change to any other class j customer, including its own, with probability p_{ij}=0.5/3, j=1, 2, 3.

Table 2 gives similar results for a large problem with 100 classes and 1000 servers. The following traffic parameters were used: ρ_i=i/1000, b_i=i, p_{i0}=0.5 and p_{ij}=0.5/100, i=1,2,..100. Table 3 gives similar results as Table 2. The assumptions are the same, with the exception that ρ_i=i/300, i=1,2,..100. We observe that the approximation model match the simulation results quite well. Some deviations were observed when the blocking probabilities are high (see for instance, Table 1, C=20, approximate and simulation results for class 3).

As mentioned above, our algorithm runs very fast. For instance, the approximation results given in Table 2 were obtained in 0.363709 seconds in Matlab 7.0.4. However, the simulation needs much longer time. The simulation results in Table 2 and 3 required around 30000 seconds. (The simulation was implemented in C program on a Pentium(R) 4 CPU 3GHz PC).

Table 1. Approximation and simulation results for 3 classes customers

	Approximation	Approximation	Approximation	Simulation	Simulation	Simulation
Capacity	class-1	class-2	class-3	class-1	class-2	class-3
C=20	0.1129	0.2252	0.3347	0.1176	0.2345	0.3485
C=25	0.0703	0.1446	0.2221	0.0722	0.1487	0.2285
C=30	0.0405	0.0856	0.1352	0.0413	0.0874	0.1380
C=35	0.0219	0.0474	0.0764	0.0225	0.0486	0.0779
C=40	0.0098	0.0217	0.0359	0.0100	0.0221	0.0366
C=45	0.0036	0.0081	0.0138	0.0037	0.0083	0.0139
C=50	0.0010	0.0025	0.0043	0.0011	0.0026	0.0045

Table 2. Approximation (Appr.) and simulation (Sim) results for 100 classes customers (1)

$class_i$	5	10	15	20	25
Appr.	0.00017	0.00035	0.00054	0.00074	0.00096
Sim	0.00016	0.00034	0.00053	0.00073	0.00094
$class_i$	30	35	40	45	50
Appr.	0.00119	0.00143	0.00169	0.00196	0.00225
Sim	0.00117	0.00140	0.00166	0.00192	0.00220
$class_i$	55	60	65	70	75
Appr.	0.00255	0.00287	0.00321	0.00357	0.00395
Sim	0.00250	0.00281	0.00315	0.00350	0.00387
$class_i$	80	85	90	95	100
Appr.	0.00435	0.00477	0.00522	0.00569	0.00618
Sim	0.00426	0.00468	0.00512	0.00558	0.00606

Next we consider the case where all customers arriving at the loss queue have the same bandwidth profile. Specifically, new customers arrive at the loss queue as class 1 and require 1 server. After an exponentially distributed service time with mean $1/\mu$, a class-1 customer changes to a class-2 customer with a bandwidth requirement of 2 servers with probability p_{12}=1. After another exponentially distributed service time

Table 3. Approximation (Appr.) and simulation (Sim) results for 100 classes customers (2)

$class_i$	5	10	15	20	25
Appr.	0.02635	0.05229	0.07780	0.10290	0.12758
Sim	0.02891	0.05751	0.08573	0.11291	0.14021
$class_i$	30	35	40	45	50
Appr.	0.15184	0.17568	0.19909	0.22208	0.24465
Sim	0.17061	0.19782	0.22293	0.24923	0.27341
$class_i$	55	60	65	70	75
Appr.	0.26679	0.28852	0.30982	0.33069	0.35115
Sim	0.30208	0.32619	0.35060	0.37237	0.39684
$class_i$	80	85	90	95	100
Appr.	0.37118	0.39081	0.41001	0.42880	0.44718
Sim	0.42342	0.44494	0.46745	0.49198	0.50767

with mean $1/\mu$, the class-2 customer changes to a class-3 customer with a bandwidth requirement of 3 servers with probability p_{23}=1. Finally, the class-3 customer departs with probability p_{30}=1 after an exponentially distributed service time with mean $1/\mu$. The offered loads are $\rho_1 > 0$, ρ_2=0, ρ_3=0, i.e., no external class-2 and class-3 arrivals occur. Given this load profile, we compare the following three bandwidth allocation strategies.
Case 1 (variable-demand policy): bandwidth is allocated on demand whenever a customer changes a class. In this case, a customer may be blocked upon arrival to the loss queue as class-1 customer and each time it changes a class. The class-dependent mean service time is $1/\mu$.
Case 2 (maximum service policy): A class 1 customer is allocated the maximum number of servers, i.e., 3 servers, upon arrival as class 1 customer to the loss queue. The mean service time is : (a). $2/\mu$ in order to keep the product of bandwidth and service-time the same as case 1; or (b). $3/\mu$ so that the arrival will use the same mean service time as case 1. The implication in case (a) is that the customer will take full advantage of the 3 servers allocated to it. In case (b) on the other hand, we assume that the customer follows its bandwidth profile and it uses only the required number of servers. A class 1 customer is blocked if these servers are not available upon arrival.
Case 3 (minimum service policy): A class 1 customer is not allowed to change bandwidth requirements. It is allocated the minimum number of customers, i.e., 1 server for a service time $6/\mu$ in order to keep the product of bandwidth and service-time the same as cases 1 and 2. A customer is blocked if no server is available upon arrival.

Case 1 was analyzed using our approximation algorithm, whereas cases 2 and 3 were analyzed using the Erlang loss formula for single class traffic. In order to facilitate the comparison among these three cases, we calculate the average call blocking probability for the case 1 as follows:

$$B_{avg} = \sum_{r=1}^{R} \bar{\rho_r} b_r B_r / \sum_{r=1}^{R} \bar{\rho_r} b_r \qquad (16)$$

We show the results in Table 4 for various values of the class-1 offered load ρ_1 for a total capacity C=100.

Table 4. Call blocking comparison among variable-demand service, Max and Min service

ρ_1	15	16	17	18	19	20	21	22
Case1	0.0495	0.0691	0.0915	0.1126	0.1336	0.1541	0.1740	0.1932
Case2a	0.0805	0.1109	0.1430	0.1755	0.2075	0.2384	0.2680	0.2959
Case2b	0.3093	0.3472	0.3817	0.4131	0.4417	0.4678	0.4917	0.5136
Case3	0.0270	0.0539	0.0874	0.1238	0.1606	0.1963	0.2302	0.2620

We note that the variable demand policy outperforms the maximum and minimum service policies when the offered load is medium or large. This observation holds for many other similar examples (not reported here).

Finally, in Table 5 we show the minimum required number of servers for a 3-class Erlang loss queue, so that the blocking probability is less than 0.01 for all three classes. The required bandwidth for the three classes is b=[1,2,3], p_{i0}=0.5 for i=1,2,3 and p_{ij}=0.5/3, j=1,2, 3 and i=1,2,3. The external offered load ρ=[ρ_1, ρ_2, ρ_3] was varied. For each set of value of ρ, we computed the minimum required servers using equation (12) (labelled as 'Asmp') and also using our algorithm in an iterative manner as explained at the beginning of section 4 (labelled as 'Appr.').

Table 5. The optimized capacity vs. offered load (ρ) for 3-classes traffic

Offered load	Method	Capacity	Offered load	Method	Capacity
[0.14 0.29 0.43]	Appr.	15	[10,20,30]	Appr.	301
[0.14 0.29 0.43]	Asmp	13	[10,20,30]	Asmp	300
[1,2,3]	Appr.	46	[40,80,120]	Appr.	1088
[1,2,3]	Asmp	44	[40,80,120]	Asmp	1087
[3,6,9]	Appr.	107	[100,200,300]	Appr.	2629
[3,6,9]	Asmp	105	[100,200,300]	Asmp	2629

6 Conclusion

In this paper, we described a model for calculation of call blocking probabilities in a multi-rate Erlang loss queue where the customers are allowed to change their bandwidth requirements during their service. Comparisons against simulation data showed that the algorithm has a good accuracy. The model was also used to evaluate different allocation policies and capacity provisioning, and we hope to expend this work in an upcoming paper. Also we are currently extending the algorithm to the case where the bandwidth requirements of each customer in service are modified by the network manager as a function of the congestion level.

References

1. Alnowibet,K.: Ph.D Dissertaion, North Carolina State University *"Nonstationary Erlang Loss Queues and Networks"*, 2004.
2. Grassmann,W. K.: *"Is the Fact That the Emperor Wears No Clothes a Subject Worthy of Publication?"*, Interfaces, 16(1986),pp. 43-46.
3. Hampshire, R. C., Massey, W.A., Mitra D. and Wang, Q. : *"Provisioning For Bandwidth Sharing and Exchange"*, Oper. Res./Comput. Sci. Interfaces Ser., 23, Kluwer Acad. Publ., Boston, MA, 2003.
4. Kaufman, J.: *"Blocking in a Shared Resource Environment"*, IEEE Transactions On Communications, Vol. COM-29, No. 10, October 1981.
5. Kelly, F. : *"Blocking probabilities in large circuit-switched networks"*,Adv. Appl. Prob., vol. 18, pp. 473-505, 1986.
6. Nilsson, A. A., Perry, M. J., Gersht, A. and Iversen, V. B.:*" On multi-rate Erlang-B computations "*, in: Proceedings of ITC 16, 1999, pp. 1051-1060.
7. Perros, H. :*Connection-Oriented Networks: SONET/SDH, ATM, MPLS and Optical Networks*, John Wiley & Sons, England, 2005.
8. Roberts, J.W.: *"A serverice system with heterogeneous user requirements"*, In Performance of Data Communications Systems and Their Applications, pp. 423-431, 1981.
9. Ross, K.W.: *"Multiservice Loss Models for Broadband Telecommunication Networks"* ,Springer-Verlag London Limited , 1995.
10. Sarkar,D. , Jewell T. and Ramakrishnan, S.: *"Convergence in the Calculation of the Handoff Arrival Rate: A Log-Time Iterative Algorithm"*,EURASIP Journal on Wireless, 2006.
11. Washington, A.N. and Perros, H.G.: *"Call blocking probabilities in a traffic groomed tandem optical network"*, Special issue dedicated to the memory of Professor Olga Casals, Blondia and Stavrakakis (Eds.) Journal of Computer Networks, Vol 45 (2004).
12. ITU-T G.7042/Y. 1035, Link Capacity Adjustment Scheme (LCAS) for Virtual Concatenated Signals , Jan. 2006.