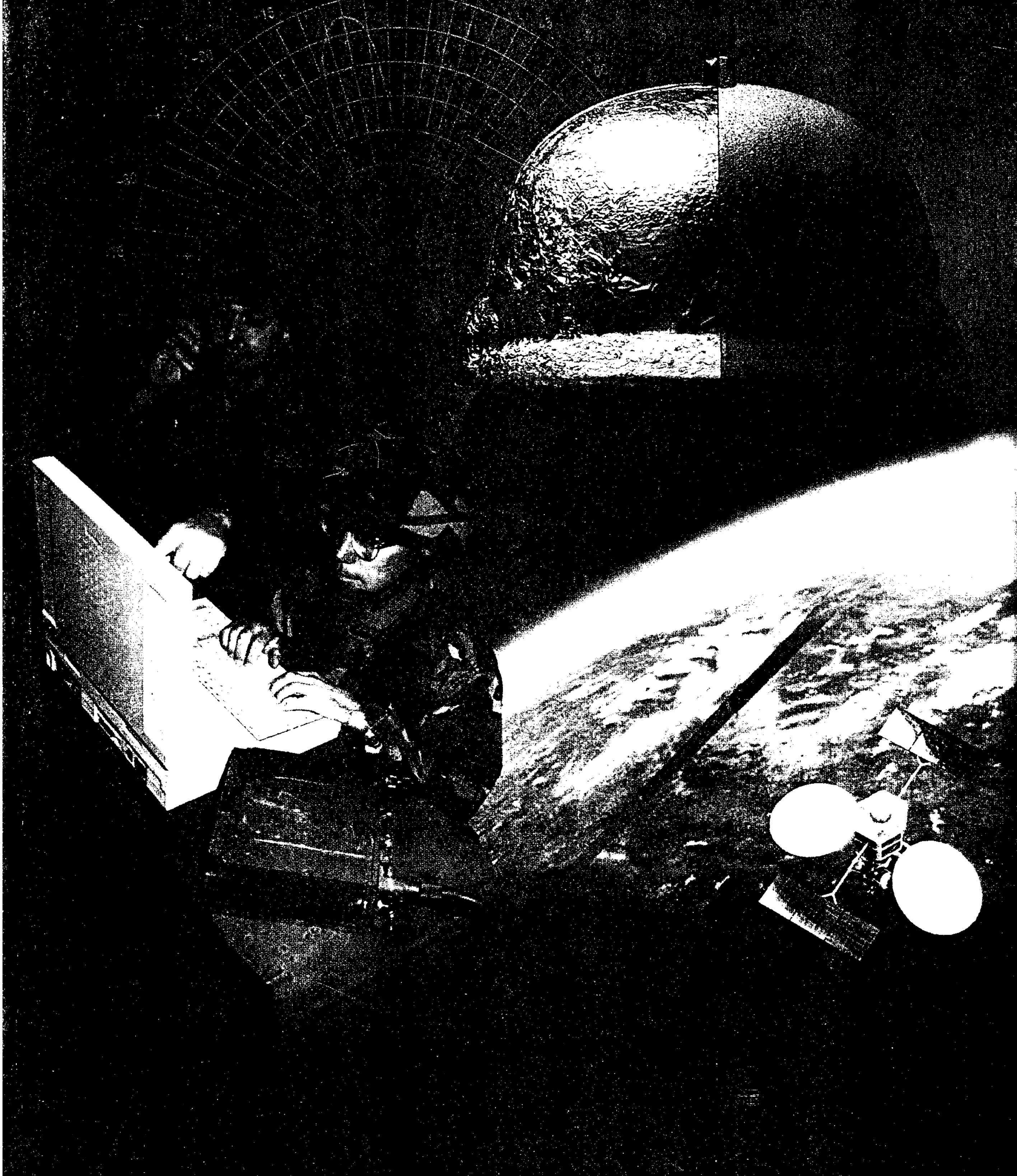
ARL FEDERATED LABORATORY



PROCEEDINGS

2ND ANNUAL FEDLAB SYMPOSIUM

FEBRUARY 2 - 6, 1998 • COLLEGE PARK, MD

Advanced Telecommunications & Information Distribution

CONSORTIUM

PERFORMANCE EVALUATION IN MULTI-RATE, MULTI-HOP COMMUNICATION NETWORKS WITH ADAPTIVE ROUTING *

Mingyan Liu, John S. Baras, Archan Misra
Department of Electrical Engineering
and Institute for Systems Research
University of Maryland at College Park
College Park, MD 20742

ABSTRACT

Accurate performance evaluation has always been an important issue in network design and analysis. Discrete event simulation has been known to be accurate but very time consuming. Thus analytical methods/approximation is necessary for real time estimation, large scale network optimization and sensitivity analysis. In a circuit-switched loss network, a particular performance metric of interest is the end-to-end blocking probability. Various analytical approaches and approximation schemes have been suggested for this problem and among them, the fixed-point method, or reduced load method, has received much attention. However, most of these schemes considered either only single traffic rate situation or multi-rate traffic under fixed routing. We develop an approximation scheme to estimate end-to-end blocking probability in a multi-rate multi-hop network with an adaptive routing scheme. The approximation results are compared with that of discrete event simulation. An example of application is also provided in which the proposed scheme is linked to the optimization tool CONSOL-OPTCAD to get network design trade-offs. The method described here is readily applicable to the accurate performance evaluation of military networks which are often large and hybrid.

INTRODUCTION

This paper is focused on the evaluation of one network performance metric – the call blocking probability of a loss network [1].

There are two types of approaches to evaluate the call blocking probabilities, or to evaluate other performance metrics. One is some sort of analytical approach (approximation, estimation) and the other is discrete event simulation. Discrete event simulation is a widely used and very helpful tool of accurate performance evaluation. However, even for a fairly small network model, it takes considerable amount of run time to get satisfactory results. With discrete event simulation, it is very hard to get the sensitivity analysis of the performance and optimization schemes, for which numerous simulation runs are needed. Military networks are typical hybrid networks, containing satellites and terrestrial wireless subnetworks. These networks are also typically large. Quite often we need to estimate the performance of a network prior to acquisition. This cannot be done in a reasonable amount of time by discrete event simulation. Another common problem is to estimate the performance of a small network within a larger network. One can do a discrete event simulation of the smaller network, while some analytical way is needed to represent the effects of the coupling and of the larger network in an aggregate fashion. For instance, the small network can be two wireless LANs serving two platoons of soldiers, while the larger network may involve many more nodes and satellites and UAVs.

Therefore this makes analytical approaches essential, since they are generally much faster than their simulation counterpart. In 1917 the Danish mathematician A. K. Erlang published his famous formula [2] which estimates the loss probability of a conventional telephone network. It covers the case of a single link with calls at a single rate. Analytically, when there are multiple links and multiple call rates, with different bandwidth requirements and with a fixed route associated with a certain source-destination pair, a loss network is modeled as a multidimensional Markov process with the dimension of the state space of the process being the product of the number of routes allowed in the network and the number of service classes with different bandwidth requirements. When alternative routes are

*Prepared through collaborative participation in the Advanced Telecommunications/Information Distribution Research Program (ATIRP) Consortium sponsored by the U.S. Army Research Laboratory under the Federal Research Laboratory Program, cooperative agreement DAAL01-96-0002 and under NASA cooperative agreement NCC3-528.

which comprises of N links expressed by $\mathcal{R}_{k,l}^{(m)} = \{(i_1,j_1),(i_2,j_2),...,(i_N,j_N)\}$. The route is in a state of admitting a call of bandwidth requirement b_s only if $C^{free}_{i_n,j_n} \geq b_s$, $n = 1,2,...,N$ under no trunk reservation admission control or $C^{free}_{i_n,j_n} \geq b_s + r_s$, $n = 1,2,...,N$ under trunk reservation admission control, where r_s is the trunk reservation parameter of the call.

Consider a route $\mathcal{R}_{k,l}^{(m)}$ which is presently available for a call, the **most congested link on the route** is the link with the fewest free circuits on this route: $\mathcal{L}_{k,l}^{(m)} = argmin_{(i_n,j_n)\in\mathcal{R}_{k,l}^{(m)}} C^{free}_{i_n,j_n}$. When there are more than one route available in the route list $\mathcal{R}_{k,l}$, the one with the maximum free bandwidth on its most congested link is selected for connecting the call. In the proposed approximation, we consider steady state, $C^{free}_{\mathcal{L}_{k,l}^{(m)}}$ must be replaced by $E[C^{free}_{\mathcal{L}_{k,l}^{(m)}}]$, $\mathcal{L}_{k,l}^{(m)}$ becomes the statistically most congested link as:

$$\mathcal{L}_{k,l}^{(m)} = argmax_{(i,j)\in\mathcal{R}_{k,l}^{(m)}} z_{i,j}$$

where $z_{i,j}$ is defined as the link load of link (i,j): $z_{i,j} = \sum_s \frac{\nu_{i,j}(s)}{\mu_{i,j}(s)} \frac{b_s}{C_{i,j}}$, where $\nu_{i,j}(s)$ is the reduced load/arrival rate for class-s calls on link (i,j). If none of the routes are admissible, the call is blocked.

Algorithm
In the proposed method, the fixed point is achieved by mappings between the following four sets of unknown variable: $\nu_{i,j}(s)$, $a_{i,j}(s)$, $p_{i,j}(n)$ and $q^{(m)}(k,l,s)$, as illustrated in the figure below:

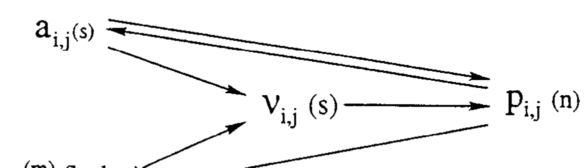


Figure 1: Relationship between variables

where $a_{i,j}(s)$ is the probability that link i,j is in a state of admitting class-s calls; $p_{i,j}(n)$ is the link occupancy probabilities for link (i,j), i.e., the probability that exactly n units of circuits are being occupied on link (i,j); and $q^{(m)}(k,l,s)$ is the probability that a call of class-s with source-destination pair of (k,l) is attempted on its m^{th} route.

Mapping: $a_{i,j}(s), q^{(m)}(k,l,s) \Longrightarrow \nu_{i,j}(s)$ The link arrival rates on link (i,j) from the node pair (k,l) for class-s traffic on the m^{th} route is given by the reduced load approximation as

$$\nu_{i,j}^{(m)}(k,l,s) = \lambda_{k,l}(s) q^{(m)}(k,l,s) I[(i,j)\in\mathcal{R}_{k,l}^{(m)}] \prod_{(u,v)\in\mathcal{R}_{k,l}^{(m)},(u,v)\neq(i,j)} a_{u,v}(s)$$

where I is the indicator function. The aggregate arrival rate of calls of class s on link (i,j) is given by

$$\nu_{i,j}(s) = \sum_{k,l}\sum_{m} \nu_{i,j}^{(m)}(k,l,s)$$

Mapping: $\nu_{i,j}(s) \Longrightarrow a_{i,j}(s), p_{i,j}(n)$
Given $\nu_{i,j}(s)$, we can compute the link occupancy probabilities $p_{i,j}(n)$ for each link in the network. This can be done by either using Kaufman's simple recursion [7] when there is no trunk reservation present, or using approaches proposed by [6, 8] as suggested by Greenberg in [4].

Mapping: $p_{i,j}(n) \Longrightarrow q^{(m)}(k,l,s)$
Given $p_{i,j}(n)$, define for link (i,j), the probability of no more than n trunks are free (at most n trunks are free) as: $\mathcal{T}_{i,j}(n) = \sum_{t=0}^{n} p_{i,j}(C_{i,j} - t)$. So the probability of attempting the call on the m^{th} route is:

$$q^{(m)}(k,l,s) = \sum_{n=0}^{C_{\mathcal{L}_{k,l}^{(m)}}} p_{\mathcal{L}_{k,l}^{(m)}}(C_{\mathcal{L}_{k,l}^{(m)}} - n) \prod_{t=1}^{m-1} \mathcal{T}_{\mathcal{L}_{k,l}^{(t)}}(n-1) \prod_{t=m+1}^{M} \mathcal{T}_{\mathcal{L}_{k,l}^{(t)}}(n).$$

End-to-end Blocking Probabilities
Finally, the end-to-end blocking probability for calls of class s between source-destination node pair (k,l) is given by

$$B(k,l,s) = 1 - \sum_{m} q^{(m)}(k,l,s) \prod_{(u,v)\in\mathcal{R}_{k,l}^{(m)}} a_{u,v}(s).$$

Repeated substitution should be used to obtain the equilibrium fixed point. And finally the end-to-end blocking probabilities can be calculated from the fixed point.

EXPERIMENTS AND EVALUATION

This example is borrowed from [4] with minor changes. The topology is derived from an existing commercial network and is depicted in Figure 2 below. The discrete event simulation was done by using OPtimized Network Engineering Tools (OPNET).
There are 16 nodes and 31 links, with link capacity ranging from 60 to 180 trunks. The detailed link-by-link traffic statistics and link capacities can be found in either [4] or [10]. The traffic in the network consists

present in addition to fixed routes, the Markov process no longer bears a product form, and the equilibrium state probabilities can be obtained by writing out the whole set of detailed balance equations and solving it [2]. However, this approach is not practical in dealing with large networks with hundreds of thousands of routes and integrated services with multiple service rates, since the computational complexity is both exponential in the number of routes and exponential in the number of service classes. This leads to the need for the development of computational techniques that provide accurate estimates of blocking probabilities for loss networks.

The reduced load approximation, also called the Erlang fixed-point method, has been proposed for this scenario and has received much attention [2]. The reduced load approximation is based on two assumptions: One is the **link independence assumption** which assumes that the blocking occurs independently from link to link. The other assumption is the **Poisson assumption** which assumes that the traffic flow to each individual link is Poisson and that the corresponding traffic rate is the original external offered rate thinned by blocking on other links of the path, thus called the reduced load.

Most of the earlier works in fixed-point method either studied the multi-rate situation with fixed routing [5], or focused on state-dependent routing schemes with single traffic rate [5], or multi-rate service with single link (resource) [6]. In [4], a fixed-point method was proposed to approximate blocking probabilities in a multi-rate multi-hop network but introduced additional computational efforts in solving the associated network reliability problem are needed, and also, the routing scheme used was sequential routing with trunk reservation, which is not often practiced in real networks.

We focus our attention on the evolving integrated service networks which bear the following characteristics: (1) The networks are typically much sparser and have a more hierarchical topology. Thus, the assumption of the existence of a direct route between source and destination nodes does not hold in most instances. (2) Routes can comprise a much larger number of hops (typically around 5 or 6) and there are typically a large number of possible routes between source and destination nodes. (3) The presence of different traffic classes characterized by widely varying bandwidth requirements and different mean holding times must be considered.

Motivated by the above, we propose to use adaptive routing in combination with the fixed-point method to calculate call blocking probabilities.

A FIXED-POINT METHOD FOR BLOCKING PROBABILITY ESTIMATION

Consider a network with J links, each indexed as (i,j), with i,j being the index of the link end nodes. Link (i,j) has a capacity of $C_{i,j}$ units of circuits or trunks. The network supports a total of S classes of connections, where a class-s connection has a bandwidth requirement of $b_s \in \mathcal{Z}^+$ on every link on the path the connection is routed. Every class-s connection also has an arrival rate of $\lambda(s)$, and the mean time of call duration is $1/\mu(s)$.

For each source-destination node pair (k,l), there is an associated set of routes $\mathcal{R}_{k,l} = \{\mathcal{R}_{k,l}^{(1)}, \mathcal{R}_{k,l}^{(2)}, ..., \mathcal{R}_{k,l}^{(M)}\}$, with each route being a subset of the set of links: $\mathcal{R}_{k,l}^{(m)} \subseteq \{1, ..., J\}$. A connection is accepted if some route has available bandwidth on each of its links to accommodate this connection, and the connection is routed on that route and holds the bandwidth for a duration with mean time $1/\mu(s)$. If none of the routes are available, the connection is rejected. The end-to-end blocking probability of a class-s connection with source destination pair (k,l) is denoted by $B_{k,l}(s)$. Throughout this paper the link is considered to be duplex and bi-directional.

The common routing policies which have been studied are fixed routing, alternative routing, sequential alternative routing and adaptive alternative routing. We focus on the last. One important scheme of this kind is called the **Least Loaded Routing** (LLR), where the call is first tried on the direct route, if there is one. If it cannot be setup along the direct route, the two-link alternative route with the largest number of point-to-point free circuits is chosen. A version of LLR was recently implemented in the AT&T long-distance domestic network [2]. The method we investigated here is an extension of LLR and is also motivated by [3] and [4]. It is a min-max scheme in that it tries to maximize the free bandwidth on the link which has the minimum free bandwidth on a route. Call admission control of the trunk reservation type is also considered.

Maximal Residual Capacity Adaptive Routing Scheme

The routing policy considered here is described as follows. Each source-destination node pair (k,l) is given a list of alternative routes $\mathcal{R}_{k,l} = \{\mathcal{R}_{k,l}^{(m)}\}, m = 1, 2, ..., M$. When a call arrives, each of the routes on the list is evaluated to determine the number of free circuits on each link employed by the routes.

Let $C_{i,j}^{free}$ denote the free/available bandwidth on link (i,j) when the call arrives, and consider a route $\mathcal{R}_{k,l}^{(m)}$

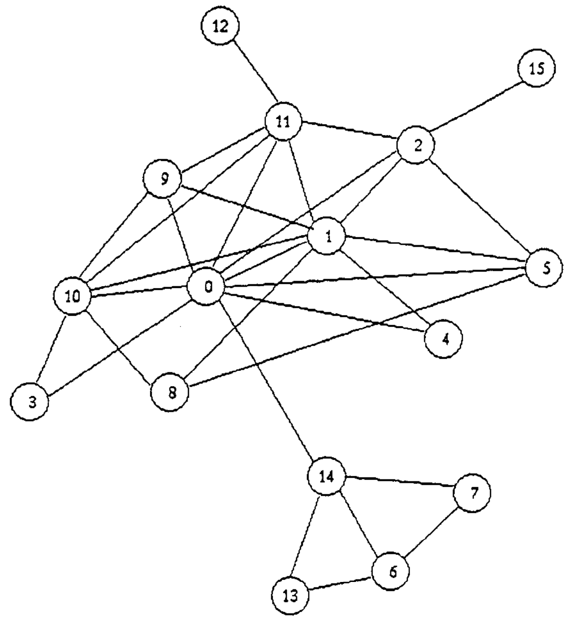


Figure 2: Topology of Example Network

of four types, namely class-1, 2, 3, and 4, and require bandwidth of 1, 2, 3, and 4 trunks, respectively. No admission control is employed.

In routing, any node pair is allowed routes that have at most 4 hops. Multiple routes for one node pair are listed in order of increasing hops, with ties broken randomly. Each link is considered to have same unit length, so only the hop number is counted.

Results for some selected node pairs and classes are listed in Table 1 through 3(FPA stands for fixed-point approximation and DES stands for discrete event simulation), each corresponding to a different traffic load. Table 1 corresponds to the "nominal" traffic which is provided in [4] and [10]. Tables 2, 3 show the results for traffic 1.4 and 1.8 times the nominal traffic, respectively.

The proposed fixed-point approximation gives conservative estimations generally, and it improves as the load gets heavier. These results strengthen the argument that these approximations are indeed very useful as estimators of worst case performance.

Node Pair	Class	FPA	DES
$(0,4)$	4	0.000178	(0.0, 0.0)
$(0,13)$	1	0.006341	(0.0021, 0.0034)
$(1,6)$	1	0.006473	(0.0030, 0.0034)
$(5,6)$	3	0.020463	(0.0189, 0.0201)
$(6,10)$	2	0.013222	(0.0109, 0.0138)
$(9,13)$	4	0.028468	(0.0185, 0.0245)
Number of Iterations		18	
CPU Time(seconds)		94.1	3.7×10^4

Table 1: Nominal Traffic.

Node Pair	Class	FPA	DES
$(0,4)$	4	0.018213	(0.0122, 0.0179)
$(0,13)$	1	0.074434	(0.0729, 0.0766)
$(1,6)$	1	0.077371	(0.0697, 0.0701)
$(5,6)$	3	0.229528	(0.2262, 0.2278)
$(6,10)$	2	0.147436	(0.1420, 0.1483)
$(9,13)$	4	0.307191	(0.2794, 0.2848)
Number of Iterations		28	
CPU Time(seconds)		145.43	4.3×10^4

Table 2: 1.4 Times The Nominal Traffic.

Node Pair	Class	FPA	DES
$(0,4)$	4	0.112658	(0.0025, 0.0026)
$(0,13)$	1	0.135564	(0.1492, 0.1500)
$(1,6)$	1	0.156322	(0.1445, 0.1466)
$(5,6)$	3	0.419781	(0.3922, 0.3940)
$(6,10)$	2	0.269145	(0.2572, 0.2583)
$(9,13)$	4	0.519083	(0.4791, 0.4793)
Number of Iterations		24	
CPU Time(seconds)		125.11	2.3×10^6

Table 3: 1.8 Times The Nominal Traffic.

APPLICATION IN NETWORK DESIGN USING CONSOL-OPTCAD

One of the main reasons we are interested in developing an analytical approximation algorithm is because such an algorithm can be easily linked to mathematical programming tools to get network performance optimization and trade-off analysis. In this section, we link the proposed reduced load approximation method with CONSOL-OPTCAD [9], which is a tool for optimization-based design, and show how network parameter design can be realized.

The application we have here is to design the trunk reservation parameters. As a way of call admission control, trunk reservation regulates individual classes of traffic as well as their interrelationship. How to choose the combination of $\{r_1, r_2, ..., r_S\}$, where r_s is the trunk reservation parameter of class-s traffic and S is the total number of different classes the network carries, to get the best network performance is important. Here the network performance of interest is the average blocking probability of the network as a whole and the blocking probabilities that each individual class of traffic experiences.

This design problem is formulated as follows: Design parameters: $r_1, r_2, ..., r_S$. Objective:

$$Minimize \ \frac{\sum_{(k,l),s} \lambda_{k,l}(s) \cdot [1 - B_{k,l}(s)] \cdot B_{k,l}(s)}{\sum_{(k,l),s} \lambda_{k,l}(s) \cdot [1 - B_{k,l}(s)]}$$

Constraints:

$$B_{k,l}(s) < bound_{k,l}(s), \quad all\ (k,l)\ and\ s$$

where $B_{k,l}(s)$ is the blocking probability of class-s traffic between source-destination node pair (k,l), and $bound_{k,l}(s)$ is the upper bound for this probability.

By applying this model to a fully connected five-node network serving 3 classes of traffic [10] with trunk reservation admission control, we get the trade-off between the blocking probability of each class vs. the weighted average blocking probability, which is shown in Table 4 and Figure 3 below (The uppermost, middle and bottom curves are blocking probabilities of Class 3, 2 and 1, respectively).

As we see from Table 4, the weighted average blocking probability and the blocking probability of class-1 type of traffic achieve their optimum at the same time with trunk reservation parameter choice of 1,4 and 5. The reason is obvious: since class-1 has much smaller trunk reservation requirement than class-2 and 3, together with its lowest bandwidth requirement, it has the highest priority and chances of being admitted into the network. On the other hand, class-2 and 3 are being jeopardized by their high trunk reservation requirement and also high bandwidth requirement. This phenomenon can also be observed in Figure 3. The curves seem very random here because they are basically the connection of discrete samples at different trunk reservation parameter combination choices.

Average	B_1	B_2	B_3	(r_1, r_2, r_3)
0.266	**0.036**	0.580	0.803	(1,4,5)
0.273	0.063	0.527	0.841	(1,3,5)
0.290	0.068	0.563	0.792	(2,4,5)
0.293	0.112	0.467	0.861	(1,2,5)
0.307	0.117	0.486	0.819	(1,2,4)
0.310	0.120	0.492	0.824	(2,3,5)
0.420	0.571	**0.284**	0.565	(4,1,2)
0.406	0.319	0.672	**0.323**	(3,5,1)

Table 4: Trunk Reservation Parameter Design

CONCLUSION[1]

In this paper we presented an approximation scheme of calculating the end-to-end, class-by-class blocking probability of a loss network with multi-rate traffic and adaptive routing scheme. It provides fairly good estimates of call blocking probabilities under normal and heavy traffic. We also presented applications, where the proposed scheme is linked to the system design optimization tool CONSOL-OPTCAD to get network parameter trade-off analysis.

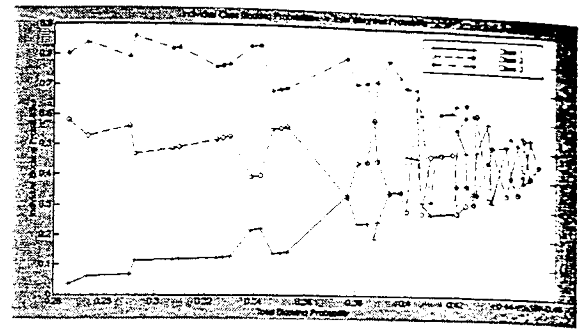


Figure 3: Trade-off in trunk reservation parameter design.

[1]The views and the conclusions expressed in this paper are those of the authors and should not be interpreted as representing the official policies, either expressed or implied of the Army Research Laboratory or the U.S. Government.

REFERENCE

[1] F. P. Kelly. Loss Networks. *The Annals of Applied Probability*, 1(3):319-378, 1991.

[2] A. Kashper, S. Chung and K. W. Ross. Computing Approximate Blocking Probabilities for Large Loss Networks with State-Dependent Routing. *IEEE Trans. Networking*, 1(1):105-115, 1993.

[3] A. Girard and M. A. Bell. Blocking evaluation for networks with residual capacity adaptive routing. *IEEE Trans. Commun.*, 37:1372-1380, 1990.

[4] A. G. Greenberg and R. Srikant. Computational Techniques for Accurate Performance Evaluation of Multi-rate, Multi-hop Communication Networks. Preprint.

[5] S. Chung and K. W. Ross. Reduced Load Approximations for Multi-rate Loss Networks. *IEEE Trans. Commun.*, 41(8):1222–1231, 1993.

[6] N. B. Bean, R. J. Gibbens and S. Zachary. The Performance of Single Resource Loss Systems in Multiservice Networks. Preprint.

[7] J. S. Kaufman. Blocking in a Shared Resource Environment. *IEEE Trans. Commun.*, 29(8):1474-1481, 1981.

[8] N. B. Bean, R. J. Gibbens and S. Zachary. Asymptotic Analysis of Single Resource Loss Systems in Heavy Traffic, with Applications to Integrated Networks. Preprint.

[9] M. K. H. Fan, A. L. Tits, J. Zhou, L. S. Wang and J. Koninchx. CONSOLE User's Manual. Technical Report TR 87-212r2, 1990.

[10] Mingyan Liu, Performance evaluation in multi-rate multi-hop communication networks, Master's Thesis, http://www.isr.umd.edu/TechReports/ISR/1997/.