Network Capacity Allocation in Service Overlay Networks

Ngok Lam[1], Zbigniew Dziong[2], Lorne G. Mason[1]

[1] Department of Electrical & Computer Engineering, McGill University, 3480 University Street, Montreal, Quebec, Canada H3A 2A7
[2] Department of Electrical Engineering, Ecole de technologie superieure,1100 Notre-Dame Street West, Montreal, Quebec, Canada H3C 1K3
ngok.lam@mcgill.ca , zdziong@ele.etsmtl.ca , lorne.mason@mcgill.ca

Abstract. We study the capacity allocation problem in service overlay networks (SON)s with state-dependent connection routing based on revenue maximization. We formulate the dimensioning problem as one in profit maximization and propose a novel model with several new features. In particular the proposed methodology employs an efficient approximation for state dependent routing that reduces the cardinality of the problem. Moreover, the new formulation also takes into account the concept of network shadow prices in the capacity allocation process to improve the efficacy of the solution scheme.

Keywords: Service Overlay networks, SON dimensioning, capacity allocations.

Section 1: Introduction:

The key components of Service Overlay Networks, (SON)s, are the SON gateways and the interconnecting logical links that lie on the top of one or more physical links. SON gateways can be treated as routers that relay service specific data and perform control functions. SON logical links provide connectivity to the SON network through existing physical links. The SON gateways are connected to adjacent gateways by the logical links. To provide SON service, the SON provider has to purchase bandwidth and QoS guarantees from the corresponding network infrastructure owners via Service Level Agreements (SLA). It is clear that the optimum amount of capacity to be purchased from the infrastructure owners, so as to maximize the net revenue, is an important issue to be faced by the SON providers.

The literature for the network dimensioning problem is usually related to circuit switched networks such as the telephone network. We shall introduce some of the works that are related to ours. In [21], Gavish and Neuman suggested a method based on Lagrangian relaxation to allocate network capacity and assign traffic in packet switching networks, but their model assumed that the traffic is routed through a single path. Medhi and Tipper did comparisons of four different approaches in [20] to a combinatorial optimization problem that describes a multi-hour network dimensioning problem for ATM networks, but their study was also based on the assumption that traffic is routed through a single path. Instead of maximizing the net income generated from the network, both of the papers chose to compute the minimum capacity allocation costs for the networks. Duan et al [22] investigated the capacity allocation problem of the SON network in order to maximize the net income gained by the SON network. Their model was also confined to networks with single fixed routes for the traffic. Girard proposed in [6] an optimization framework for dimensioning circuit-switched networks employing a more flexible load sharing alternative routing scheme. This framework was applied in [5] for the dimensioning of telephone networks. The formulations in [5] and [6] were problem specific in that they dimension circuit-switched networks consisting of only one-link and two-link paths. Shi and Turner presented in [7] a heuristic approach to size SON multicast networks. Their main focus was on the routing algorithms that optimize the delays and the bandwidth usage on the multicast service nodes. The dimensioning uses a simple algorithm that equalizes the residual capacities across the multicast network.

In our study, we dimension the SON network based on revenue maximization. In this aspect we are not only considering the net income in the objective function, but we are also incorporating the notion of average network shadow price in the dimensioning process in order to reflect the sensitivity of net revenue to the dimensions of the links. We consider the SON network as a generic network and provide a framework for dimensioning based on the traffic rewards. The dimensioning problem is formulated as a constrained optimization problem for two distinct routing models. From the KKT conditions of the optimization formulation, we devise an iterative method that leads to near-optimal solutions. Compared with the previous studies reported in the literature, our models allow more flexible routing schemes whereas each path can be comprised of an arbitrary number of links. We also incorporate two sophisticated routing schemes to better approximate the state dependent routing scheme assumed in the SON environment. We present analytical optimization models, and include detailed discussions of the implementation issues, as well as numerical studies that verify the models' accuracy. A novelty of our study is that we provide an economic integration of the control layer and the dimensioning layer through the use of average shadow price concept.

This article is structured as follows: section 2 will be devoted to the description of routing algorithms used in later sections. The mathematical formulation is included in section 3 together with the details of the analytical models for the network dimensioning problem. We discuss implementation issues and present numerical results in section 4. The conclusion is given in section 5.

Section 2: Routing Algorithms

As mentioned above, in our SON framework we apply the state dependent reward maximization routing strategy; such as the MDPD strategy [4], in order to achieve integrated economic framework. Nevertheless, to simplify analytical performance evaluation, in our dimensioning model we approximate MDPD routing strategy by a routing based on a load sharing concept. The pure load sharing routing strategy is inefficient as calls can be lost even when valid available paths are present which is not the case with MDPD approach. To overcome this issue we employ two relatively simple yet efficient load sharing routing strategies to provide conservative approximations to the MDPD strategy. The blocking performances of these two strategies provide upper bounds for the MDPD strategy. The dimensioning solutions based on them are therefore conservative.

The first routing strategy used here is known as the "combined load sharing and alternate routing" strategy [11]. We denote this strategy as routing strategy *I* and the corresponding optimization model as model *I* throughout the article. In this routing strategy, the potential paths for a traffic flow are ordered to form a set of routing sequences. Each of these routing sequences consists of all the potential paths for the traffic flow. The paths are arranged in different orders in different routing sequences. Every routing sequence bears a load sharing coefficient; the traffic flow is assigned to a routing sequence with probability proportional to the load sharing coefficient of that sequence. The traffic flow must attempt all the paths in its assigned sequence before declaring connection failure. A connection would fail if and only if all the paths in the assigned sequence are blocked. Figure 1 shows an instance of such a scheme for the traffic flow between the nodes S and D, f_{SD}. In that example, the first routing sequence carries a fraction $a_1/(a_1+a_2+a_3)$ of the total traffic between nodes S and D, and the flow f_{SD} must attempt paths in the order *P1*, *P2*, *P3*. The second sequence in the example carries $a_2/(a_1+a_2+a_3)$ of the traffic and the paths must be attempted in the order *P2*, *P3*, *P1*. The third sequence carries $a_3/(a_1+a_2+a_3)$ of the traffic flow f_{SD}, and the paths must be attempted in the order *P3*, *P1*, *P2*.

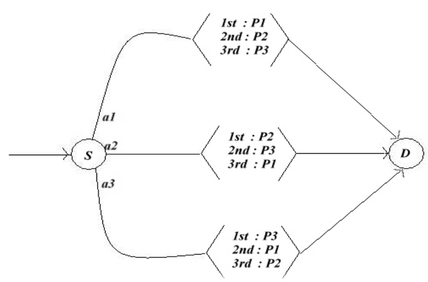

Figure 1: routing strategy I

In the second strategy considered for approximation of MDPD routing, each potential path for a particular traffic flow is assigned a routing coefficient. First, the traffic flow is assigned to a path with probability proportional to the routing coefficients. If this path is blocked, the scheme will attempt the remaining *n-1* paths with probabilities proportional to the paths' original routing coefficients. If the new path chosen by the scheme also turns out to be blocked, the traffic will attempt the remaining *n-2* paths with probabilities proportional to their original routing coefficients. This process continues until either the traffic flow is routed or until the scheme discovers that all *n* paths are blocked. Figure 2 depicts a case of such a routing scheme. In that scenario, path *P1* is discovered blocked by a traffic flow assigned to it. The traffic therefore overflows to the remaining paths *P2* and *P3*, with probabilities directly proportional to their routing coefficients a_2 and a_3. In the remainder of this paper the second strategy is referred to as routing strategy *II* and the corresponding optimization model as model *II*.

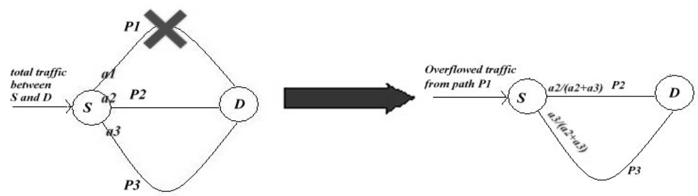

Figure 2: routing strategy II

Section 3: The Optimization Models

We treat the SON network as a generic network and the SON gateways as generic networking nodes that ship data to generate revenues. In the following discussion we shall use "SON gateways" and "nodes" inter-changeably. The traffics considered here are homogenous traffics with the same bandwidth requirement following the exponential distribution for both their inter-arrival time and service time. We assume, at this stage, that the network topologies, the traffic intensities, the GoS requirements, and traffic revenues (or service prices) are all given parameters. We also assume that there exists at most one physical link between any pair of nodes in the underlying network, although it is possible that there exist more than one physical links. The GoS requirements are specified in the form of blocking probabilities. We formulate the problem as a dimensioning problem for the links in the network. For the sake of implementation, we leave most of the partial derivatives in the equations so as to enable numerical methods such as the finite difference method to be employed.

Without loss of generality, let's assume the paths in the sequence set q are indexed by the order they will be attempted in routing scheme *I*. It is easy to see that because we assumed every routing sequence for a particular flow $f_{ij.}$ contains all the corresponding end to end paths, the blocking probability for a particular traffic flow $f_{ij.}$ under strategy *I* can be written as:

$$\sum_{q} \alpha_q^{ij} \left(\prod_{k=2}^{|R_{ij}|} P(r_k^q \quad blocked \mid r_i^q \quad blocked, i = 1, 2, \ldots k-1) \right) P(r_1^q \quad blocked) \tag{1}$$

where α_q^{ij} is the probability of selecting the sequence q for traffic pair $f_{ij.}$, r_k^q is a particular path indexed by k in the routing sequence q of the traffic pair $f_{ij.}$, R_{ij} are all the end-to-end paths for the flow $f_{ij.}$. To further simplify (*1*), we can assume independence of the paths r_k^q, $k=1,2,\ldots|R_{ij}|$. As a consequence of this assumption, all the conditional probabilities of (*1*) are reduced to the unconditional probabilities, equation (*1*) is now given by:

$$\prod_{r \in R_{ij}} B_r \tag{2}$$

where B_r is the blocking probability for a particular path r. The approximation of (*2*) is exact when none of the paths r shares a common link. For the sake of simplicity we use (*2*) in the formulation of our model. A sidebar note is that the calculation of (*1*) can be equivalently viewed as finding the probability such that at least one of the cut sets with respect to R_{ij} has all its elements failed, in our case it is possible to use a recursive technique to tackle that without the need of finding the sets explicitly, we shall have a short discussion about this in section 3.

Now let's go to derive the overflow traffic generated by routing strategy *I*. We can see the amount of traffic overflowing to a particular link *s*, as a result of routing scheme *I,* is as follows:

$$\sum_{ij}\sum_{q\in Q_{ij}}\sum_{k\geq 1, r_k\in q}((\lambda^{ij}\alpha_q^{ij}(\prod_{k>m, r_m\in q} B_{r_m})(1-B_{r_k}))(\delta_{s\in r_k})/(1-B_s)) \qquad (3)$$

where r_m are all the paths preceding r_k in the sequence q. We define r_0 to be a dummy path that does not consist of any physical links, and we artificially defined $B_{r_0}=1$ for consistency. Note that we use the same independence assumption as that of (*2*) in formulating the blocking probabilities. In the expression (*3*), $\delta_A = 1$ if event A is true, and $\delta_A = 0$ if event A is false, q refers to a particular sequence of paths, and Q_{ij} is the set containing all the sequences of paths for the flow f_{ij}. A particular traffic flow $f_{ij.}$ will overflow to the link *s* if this link is used by a path r_k which is contained in one of the sequences, *q,* inside the set Q_{ij}, and all the paths r_m before the path r_k in the sequence *q* are blocked. This results in the expression (*3*) which will be useful in calculating the link blocking probabilities for our model.

We present the optimization model for routing strategy *I* first, and denote it as model *I* throughout this article. Optimization model *II* for the routing strategy *II* is similar although the expression for overflow traffic is more complicated. Let's first define the variables being used in model *I* in the table below:

Table 1. The set of expressions used.

$C_s(N_s)$	= the cost function for having a capacity of N_s on link *s*.
w_r^{ij}	= the revenue generated by traffic flow $f_{ij.}$ (i.e. traffic from node *i* to node *j*) through path *r*.
B_r	= the blocking probability of path *r*.
λ^{ij}	= the offered traffic in terms of number of connections for the flow f_{ij}.
λ_r^{ij}	= the carried traffic for flow $f_{ij.}$ on a path *r,* it is equal to $\lambda^{ij}\Sigma_q(\alpha_q^{ij}\mathrm{P}(C_r^q))$, where α_q^{ij} is the load sharing coefficient corresponding to a sequence *q*, and $\mathrm{P}(C_r^q)$ is the probability that the traffic is being admitted at route *r* of the sequence *q,* where $q\in Q_{ij}$.
$\bar{L}^{ij}$	= the upper bound for the end-to-end blocking probability of the flow $f_{ij.}$.
α_q^{ij}	= the probability of selecting sequence *q* for the flow $f_{ij.}$.
$E(a_s,N_s)$	= the Erlang-B equation for the link s, with offered traffic a_s and capacity N_s.
R_{ij}	= the set containing all the possible end-to-end paths for the flow $f_{ij.}$.
R^s	= the average shadow price for the link *s*, this is a sensitivity measurement of the total revenue with respect to the link capacity of link *s*.

The optimization formulation is shown in equation (*4*), where x^{ij}, v^{ij}, u_q^{ij}, y_s, z_s are the KKT multipliers. The Lagrange equation for (*4*) is shown in equation (*5*).The first order KKT conditions of (*5*) are listed in equation (*6*). Equation (*6.III*) involves the term R^s, which is the sensitivity of the total revenue with respect to the link capacity of link *s* and is derived as a partial derivative of the revenue with respect to the link capacity. This term tends to be ignored in some of the literature, but we discovered that the addition of this term enables our methodology to yield better results, since it takes into account of the impact of the link capacities on the total revenue generated and reflects the knock-on effects of dimensioning link *s* over the total revenue. This term is also known as the average network shadow price [4] for the link *s.* Link shadow price is being used extensively in the routing literature as a control parameter to improve network resource utilization and therefore the incorporation of the average link shadow price in the dimensioning process forms an economic framework that integrates the dimensioning model with the control model of the SON network.

$$
\begin{aligned}
& Min\Big(\sum_s C_s(N_s) - \sum_{i,j,r} w_r^{ij}\lambda_r^{ij}\Big) \\
& s.t. \\
& \quad \prod_{r\in R_{ij}} B_r \le \overline{L}^{ij} \qquad --- (x^{ij}) \\
& \quad \sum_{q\in Q_{ij}} \alpha_q^{ij} = 1 \qquad --- (v^{ij}) \\
& \quad \alpha_q^{ij} \ge 0 \qquad --- (u_q^{ij}) \\
& \quad E(a_s, N_s) = B_s \; --- (y_s) \\
& \quad N_s \ge 0 \qquad --- (z_s)
\end{aligned}
\tag{4}
$$

$$
\begin{aligned}
L = & \sum_s C_s(N_s) - \sum_{i,j,q,r_k\in q} w_{r_k}^{ij}\lambda^{ij}\alpha_q^{ij}\Big(\Big(\prod_{k>m,\, r_m\in q} B_{r_m}\Big)(1-B_{r_k})\Big) + \sum_{ij} x^{ij}\Big(\prod_{r\in R_{ij}} B_r - \overline{L}^{ij}\Big) \\
& + \sum_{ij} v^{ij}\Big(\sum_{q\in Q_{ij}} \alpha_q^{ij} - 1\Big) - \sum_{ij}\sum_{q\in Q_{ij}} u_q^{ij}\alpha_q^{ij} + \sum_s y_s\big(E(a_s, N_s) - B_s\big) \\
& - \sum_s z_s N_s
\end{aligned}
\tag{5}
$$

Note that the KKT multiplier v^{ij} at left hand side of equation (*6.I*) is independent of the path taken, *r*. As a consequence, the expression at the right hand side of the equation should have the same value for all the paths carrying a non-zero traffic portion of the flow $f_{ij.}$. This equation can be treated as the optimality equation for the optimal routing problem. This implies that all the paths with a positive share of the traffic $f_{ij.}$ should have the same marginal cost for the flows they carry. This is a well known fact for system optimality in the literature. Because of the complementary slackness conditions, we can further conclude from (*4)* that all paths with positive shares of traffic $f_{ij.}$ should have the multiplier u_q^{ij} being 0. Together with the multipliers y_s from the previous iteration (or from the initial values) we can solve (*6.I*) for α_q^{ij}. With the routing coefficients α_q^{ij} and the multipliers x^{ij} (from the previous iteration or from the initial values) we can calculate the multipliers y_s from (*6.II*). With the multipliers y_s we can solve the optimal dimension sub-problem in (*6.III*).

$$v^{ij} = u_q^{ij} + \sum_{r_k \in q} w_{r_k}^{ij} \lambda^{ij} \Big(\prod_{k>m, r_m \in q} B_{r_m} \Big)(1 - B_{r_k}) - \frac{\partial \big(\sum_s y_s E(a_s, N_s)\big)}{\partial \alpha_q^{ij}} \qquad (I)$$

$$y_s = \sum_{i,j,q,r_k \in q} w_{r_k}^{ij} \lambda^{ij} \alpha_q^{ij} \Big(\frac{\partial \big(\prod_{k>m, r_m \in q} B_{r_m} \big)(1 - B_{r_k})}{\partial B_s} \Big) + \sum_{ij} x^{ij} \Big(\frac{\partial \big(\prod_{r \in R_{ij}} B_r\big)}{\partial B_s} \Big) + \sum_{s'} y_{s'} \frac{\partial E(a_{s'}, N_{s'})}{\partial B_s} \qquad (II)$$

$$C_s^{'}(N_s) = R^s - \frac{\partial \big(\sum_{ij} x^{ij} \big(\prod_{r \in R_{ij}} B_r \big) \big)}{\partial N_s} - y_s \frac{\partial E(a_s, N_s)}{\partial N_s} + z_s \qquad (III)$$

(6)

Then with all the multipliers and the optimization variables, we can solve the dual of (*5*) for the multipliers x^{ij} which is: $\max_x (L(x))$, where *L(x)* is the function $\min_{N,\alpha} L$ with *x* as the variable, *L(x)* is continuous and concave for any primal, but if the primal problem has non-unique solutions. *L(x)* is non-differentiable. To get around with this, we employ the sub-gradient method to maximize *L(x)*. With the new multipliers x^{ij}, we can go back to equation (*6.I*) and restart the whole process again until the solution converges. When the solution converges, that means equations (*6.I*), (*6.II*) and (*6.III*) will all be satisfied, which implies the first order KKT conditions of the optimization problem are satisfied and the solution of the dimensioning problem arrives at a stationary point. It is always possible to perform a second order optimality condition check to test for local optimality, although the computation of the Hessian matrix for the Lagrange equation (*5*) can be expensive as the size of the Hessian is of order $|S|^2$, where *S* is the set containing all the links of the network. The main solution given by this model is the dimension of the individual links. The iterative solution scheme employed here is similar to that of [6], but the formulation of [6] would be exceedingly complicated if it is restructured to suit the multi-link paths in the SON environment. Our formulation, on the other hand, can tackle paths consisting of an arbitrary number of links without making any modification. Moreover, we take into account the notion of average shadow price in dimensioning each of the individual links. This is something missing in other studies. Additionally, the more sophisticated routing schemes used here also improve the network resource utilization, which in turn helps to alleviate the problem of over-dimensioning in the final solutions.

Though routing strategy *II* appears to be more complicated in the sense that it attempts the paths to choose a route, it can be proven that the blocking performance expressions are the same as strategy *I*. Both strategies would declare connection failure for a particular traffic pair $f_{ij.}$, if and only if all the possible paths are blocked. In other words, we can use the same approximation (*2*) to represent the blocking performance for strategy *II*. The expression for overflow traffic of traffic $f_{ij.}$, however turns out to be more complicated for strategy *II*. We assume that strategy *II* only overflows unblocked paths, which is the result of maintaining an up-to-date path status table. We can show that the expression for overflow traffic to a link *s*, is represented by (*7)*. Note that in this equation, $|R_{ij}|$ is the cardinality of the set R_{ij}. Θ_k^{ij} is a set that contains some particular sets as elements - each of the elements is itself a set that contains k different paths for the traffic f_{ij}., we denote these elements by b_k^{ij} , and Θ_k^{ij} holds all the possible b_k^{ij}.

$$\sum_{ij} \sum_{k=1}^{|R_{ij}|-1} \sum_{r \notin b_k^{ij}, b_k^{ij} \in \Theta_k^{ij}} [[P(A_r)P(B_k \mid A_r)(\sum_{r'' \in b_k^{ij}} \alpha_{r''}^{ij} \lambda^{ij})(\frac{\alpha_r^{ij}}{\sum_{r' \in R_{ij}, r' \notin b_k^{ij}} \alpha_{r'}^{ij}})]\delta_{s \in r}] \tag{7}$$

A_r in the above equation denotes the event "path r is not blocked", and B_k denotes the event "only the k paths in b_k^{ij} are blocked".

$$\begin{aligned} v^{ij} = {} & u_r^{ij} + w_r^{ij} \lambda^{ij} P(A_r) + w_r^{ij} \sum_{k=1}^{|R_{ij}|-1} \sum_{r \notin b_k^{ij}, b_k^{ij} \in \Theta_k^{ij}} P(A_r)P(B_k \mid A_r) \frac{\sum_{r' \neq r, r' \notin b_k^{ij}, r' \in R_{ij}} \alpha_{r'}^{ij}}{(\sum_{r' \notin b_k^{ij}, r' \in R_{ij}} \alpha_{r'}^{ij})^2} \sum_{r'' \in b_k^{ij}} \lambda^{ij} \alpha_{r''}^{ij} \\ & + \sum_{i,j,\tilde{r} \in R_{ij}} w_{\tilde{r}}^{ij} \sum_{k=1}^{|R_{ij}|-1} \sum_{\tilde{r} \notin b_k^{ij}, b_k^{ij} \in \Theta_k^{ij}} [P(A_{\tilde{r}})P(B_k \mid A_{\tilde{r}}) \frac{\alpha_{\tilde{r}}^{ij} \times \lambda^{ij}}{\sum_{r' \notin b_k^{ij}, r' \in R_{ij}} \alpha_{r'}^{ij}}]\delta_{r \in b_k^{ij}} \\ & - \sum_s y_s \frac{\partial E(a_s, N_s)}{\partial \alpha_r^{ij}} \qquad (I) \end{aligned}$$

$$\begin{aligned} y_s = {} & \sum_{i,j,r \in R_{ij}} w_r^{ij} \lambda^{ij} \alpha_r^{ij} \frac{\partial (P(A_r))}{\partial B_s} \\ & + \sum_{i,j,r \in R_{ij}} w_r^{ij} \sum_{k=1}^{|R_{ij}|-1} \sum_{r \notin b_k^{ij}, b_k^{ij} \in \Theta_k^{ij}} \frac{\partial (P(A_r)P(B_k \mid A_r))}{\partial B_s} (\frac{\alpha_r^{ij}}{\sum_{r' \notin b_k^{ij}, r' \in R_{ij}} \alpha_{r'}^{ij}}) \sum_{r'' \in b_k^{ij}} \lambda^{ij} \alpha_{r''}^{ij} \\ & - \sum_{ij} x^{ij} \frac{\partial \prod_{r \in R_{ij}} B_r}{\partial B_s} - \sum_s y_s \frac{\partial E(a_s, N_s)}{\partial B_s} \qquad (II) \end{aligned}$$

$$C_s^{'}(N_s) = R^s - \frac{\partial (\sum_{ij} x^{ij} (\prod_{r \in R_{ij}} B_r))}{\partial N_s}{}_s - y_s \frac{\partial E(a_s, N_s)}{\partial N_s} + z_s \qquad (III)$$

$$(8)$$

The optimization model for routing strategy *II* is similar to that of strategy *I* although the equations are more complicated because of the overflow pattern. To save space, we shall only list the set of first order KKT conditions in (*8*). Again R^s is the sensitivity of the total revenue with respect to the link capacity of link *s*. The above KKT conditions can be solved by using an iterative approach as was done in model *I*.

Section 4: Numerical Results and Discussions

We conducted a series of numerical studies with the mathematical models. We feel it is useful to give a brief discussion of some of implementation issues of both models. In our implementations we use the Frank-Wolfe method to compute the load sharing coefficients. As the Frank-Wolfe method may converge very slowly when it is close to the optimal solution; we artificially supply an upper bound for the number of iterations. This slightly decreases the accuracy of the solution, but in general the efficiency of Frank-Wolfe method is improved.

We use the Erlang B formula extensively in the models here. Direct implementation of the Erlang B formula suffers from two major problems. First, the magnitude of its components explode with the capacity and the offered traffic. Second, direct calculation of the Erlang formula would require a time complexity of O(n^2) where n is the capacity of the link. Both problems can be circumvented by the method mentioned in [18], and the time complexity is reduced to O(n) in our implementation. One further difficulty related to the original Erlang formula is that it is a discrete function in the capacity. A continuous version of Erlang B equation available in the literature[8] involves complicated components that make the computation inefficient. We take advantage of the fact that the Erlang B formula is a strictly decreasing function in the capacity, and use the linear interpolation method to approximate the continuous valued capacity.

For model *II*, a great deal of the difficulty lies in the computation of the expression $P(B_k|A_r)$. A recursive style algorithm can be used to calculate the overflow traffic elegantly. We employ a recursive DFS (Depth First Search) algorithm [19] to search through all combinations of path failures that can result in k path failures so as to avoid the complexities involved in calculating the sets Θ_k^{ij} explicitly. The same code can also be exploited to find all the cut sets between node i and node j.

Preliminary numerical studies were conducted on two relatively small sample networks for both of the models. The first sample network is illustrated in figure 5. For this network, we assume there are two pairs of traffic, one is from node A to node B, with an average connection rate of 6 units and revenue of 7 units for each carried connection, and the other traffic is from node A to node C, with an average connection rate of 5 units and revenue of 8 units per connection. The GoS requirement is 0.1 for both traffic demands.

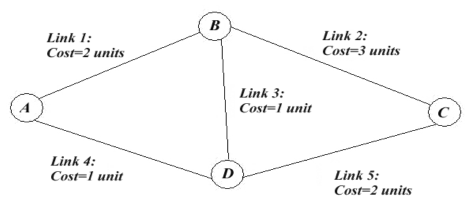

Figure 5

The following table summarizes results obtained by the two models, the results are rounded to integers:

Table 2. Results for the sample network in figure 5.

Link index	*1*	*2*	*3*	*4*	*5*	Cost	Net reward
Model I	0	0	13	20	9	51	28
Model II	0	0	12	20	9	50	29

The GoS constraints are satisfied by both assignments. Model *II* generates slightly more net reward than model *I*. A plausible explanation to this phenomenon is that in general the routing strategy *I* can generate different load distribution on the considered path when compared with the routing strategy *II*. This fact, combined with significantly larger number of variables to optimize in case of strategy *I,* may lead to a more suboptimal solution in model *I*. The convergence graphs for both models are shown in figure 6 below, the y-axis corresponds to the net revenue, while the x-axis corresponds to the iteration number. As we can see, both of them converge in approximately 5 iterations, and each iteration takes around 6 seconds of time on a 1.4Ghz P4 machine for the both models.

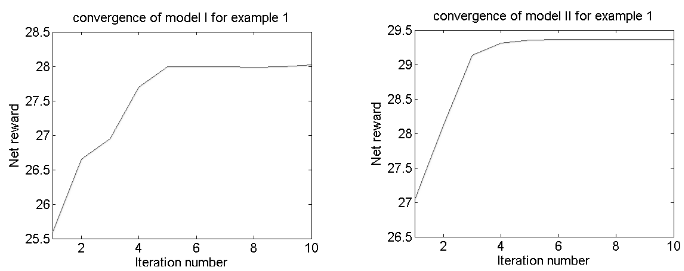

Figure 6

We also considered a larger problem as shown in figure 7. This problem has 5 pairs of traffic flows, the traffic details are listed in table 3 and the dimensioning results are depicted in table 4.

Table 3. Traffic matrix for the network in figure 7.

	Average Rate	Revenue per connection	Possible routes (indexes of the links)	GoS
Traffic A-> B	25 units	18 units	1 4->5->2	0.1
Traffic B->A	15 units	12 units	1	0.1
Traffic A->C	18 units	25 units	1->2 4->5 4->7->6	0.1
Traffic B->C	30 units	17 units	2 1->4->5	0.1
Traffic E->D	12 units	18 units	7->8 5->3 7->6->3	0.1

Table 4. Dimensioning results for the network in figure 7.

Link index	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	Cost	Net reward
Model I	74	51	9	0	0	24	30	13	563	1209
Model II	75	55	0	0	0	0	30	16	524	1227

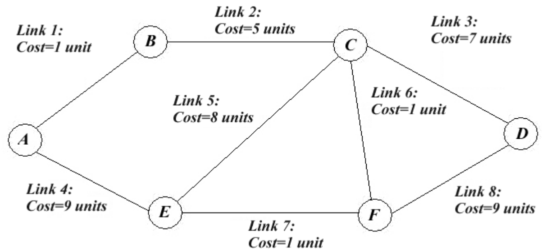

Figure 7

It takes approximately 30 iterations for both models to converge in this larger network, with each iteration taking approximately 8 seconds. Model *II* again gives higher net revenue while satisfying all the GoS requirements. This is to be expected as according to our performance models, model *II* generates less overflow traffic. As a result, less resource is needed for model *II* to meet the GoS Constraints, which becomes more trivial in this example. Depending on the implementation, model *I* might further suffer from the problem of large cardinality in generating sequences, as the number of sequence grows as a factorial function of the possible paths. One may have to limit the number of sequences generated by model *I* in large examples and this could be another disadvantage of model *I* for large-size real-world networks,

Section 5: Conclusions

We studied the problem of SON dimensioning problem by employing an iterative process based on two different routing models. A major contribution of this study is that we provide an approach to dimension the SON network by considering the SON network as a generic network based on the traffic revenues. Moreover the concept of average link shadow price is also incorporated in the SON dimensioning models. We also provided numerical results to offer insights into the efficacy of the theoretical models. The numerical results are promising on the small sample networks tested, and convergence usually occurs within a few iterations. The current effort is to capture the key features of the solution scheme so as to improve computational performance. The verification of the solution quality through state dependent routing simulations and the convergence studies are both in progress. Overall the study reported here provides, under the new perspective of profit maximization, an economic integration of the control and dimensioning layers when allocating capacities in the SON network.

References:

1. Zbigniew Dziong, ATM network resource management. McGraw-Hill, New York, 1997.
2. Michal Pioro and Deepankar Medhi, Routing flow and capacity design in communication and computer networks. Morgan Kaufmann Publishers, 2004.
3. Frank P. Kelly, "Routing in circuit-switched networks: Optimization, Shadow Prices and Decentralization", Advanced In Applied Probability, Vol. 20, No. 1, 1988, 112-144.
4. Zbigniew Dziong and Lorne G. Mason, "Call admission and routing in multi-service loss networks", IEEE Transactions On Communications, Vol. 42, No.2/3/4, 1994, 2011-2022.

5. Andre Girard and Bernard Liau, “Dimensioning of adaptively routed Networks”, IEEE Transactions On Networking, Vol. 1, No. 4, 1993, 460-468.
6. Andre Girard, “Revenue optimization of telecommunication networks”, IEEE Transactions On Communications, Vol. 41, No. 4, 1993, 583-591.
7. Sherlia Shi and Jonathan S. Turner, “Multicast routing and bandwidth dimensioning in overlay networks”, IEEE Journal On Selected Areas In Communications, Vol. 20, No. 8, 2002, 1444-1455.
8. R. F. Farmer and I. Kaufman, “On the numerical evaluation of some basic traffic formulae”, Networks, Vol. 8, No. 2, 1978, 153–186.
9. Dimitri Bertsekas and Robert Gallager, Data Networks. Englewood Cliffs, NJ: Prentice Hall, 1991.
10. Sheldon M. Ross, Introduction to Probability Models. Academic Press, 1993.
11. Andre Girard, Routing and dimensioning in circuit-switched networks. Addison-Wesley Publishing Company, 1990.
12. Deepankar Medhi and Sujit Guptan, ”Network dimensioning and performance of multi-service, multi-rate Loss networks with dynamic routing”, IEEE Transactions On Networking, Vol. 5, No. 6, December,1997, 944-957.
13. Mingyan Liu and John S. Baras, “Fixed Point approximation for multi-rate multi-hop Loss networks with state-dependent routing”, IEEE Transactions On Networking, Vol. 12, No.2, April, 2004, 361-374.
14. Andre Girard and Brunilde Sanso, “Multicommodity flow models, failure propagation, and reliable loss network design”, IEEE Transactions On Networking, Vol. 6, No. 1, February, 1998, 82-93.
15. Sun-Ping Chung and Keith W. Ross, “Reduced load approximations for multi-rate loss networks”, IEEE Transactions On Communications, Vol. 41, No. 8, August 1993, 1222-1231.
16. Syed I. A. Shah and Andre Girard, “Multi-Service network design: a decomposition approach”, Proc. Globecom 98, 3080-3085.
17. Albert G. Greenberg and R. Srikant, “Computational techniques for accurate performance evaluation of multi-rate, multi-hop communication networks”, IEEE Transactions On Networking, Vol. 5, No. 2, April 1997, 266-277.
18. SanZheng Qiao and Liyuan Qiao, “A robust and efficient algorithm for evaluating Erlang B formula”, Technical Report CAS98-03, Department of Computing and Software, McMaster University, 1998.
19. T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, 2nd edition. The MIT Press, 2001.
20. D. Medhi and D. Tipper, “Some approaches to solving a multi-hour broadband network capacity design problem with single-path routing”, Telecommunication Systems, Vol. 13, No. 2, 2000, 269–291.
21. B. Gavish and I. Neuman, “A system for routing and capacity assignment in computer communication networks”. IEEE Transactions On Communications, Vol. 37, 1989, 360-366.
22. Zhenhai Duan, Zhi-Li Zhang, and Yiwei Thomas, “Service Overlay Networks: SLAs, QoS, and bandwidth provisioning”, IEEE/ACM Transactions On Networking, Vol.11, No. 6, December 2003, 870-883.