Capacity Allocation in Service Overlay Networks - Maximum Profit vs Minimum Cost Formulations

Ngok Lam
Department of Electrical and Computer Engineering, McGill University
3480 University Street,
Montreal, Quebec, Canada H3A 2A7
ngok.lam@mcgill.ca

Zbigniew Dziong
Department of Electrical Engineering, Ecole de Technologie Superieure
1100 Notre-Dame Street West,
Montreal, Quebec, Canada H3C 1k3
zdziong@ele.etsmtl.ca

Lorne G. Mason
Department of Electrical and Computer Engineering, McGill University
3480 University Stree
Montreal, Quebec, Canada H3A 2A7
lorne.mason@mcgill.ca

ABSTRACT

We studied a class of Service Overlay Network (SON) capacity allocation problem with Grade of Service (GoS) constraints. Similar problems in the literature are typically formulated as either a Maximum Profit (MP) optimization problem or a Minimum Cost (MC) optimization problem. In this article we investigate the relationship between the MP and MC formulations.

When the service charges are zero, the MP and MC formulations are obviously equivalent. By using the set of Lagrange multipliers from the MC formulation as a tool, we investigate the extent that this equivalence holds with respect to the service charges. The same set of multipliers also acts as thresholds for service charges so that the SON operator will be happy to provide adequate service level even if he/she is not obligated to do so. A key contribution of this paper is the provision of insight into the solution nature of the MP and the MC formulations under different service charge parameters, thereby giving guidelines to the proper formulation should be chosen. The second contribution is the use of the Lagrange multipliers in providing pricing information to the network operators.

Keywords

Service overlay network design, GoS, Capacity allocation, Profit maximization, Cost minimization, Optimization methods, Pricing of network services

1. INTRODUCTION

It is not uncommon that in formulating network design problems, one may be confronted with the choice of choosing whether to minimize the investment or to maximize the profit from the network. While this choice of the criterion of optimization could have profound influence on the final solution obtained, we note that in the literature the choices are usually made in a somewhat undisclosed manner without giving the detailed background rationale to the readers. A natural question is: Since the MC and MP formulations optimize different objectives, how do their solutions compare? Or to be specific, for a given set of parameters, which formulation should be picked and how do the solutions compare. This paper addresses the objective function choice problem faced by the operator for a class of Service Overlay Network (SON) capacity allocation problems. We identify the region such that Maximum Profit (MP) and Minimum Cost (MC) formulations offer the same solution. We also identify the region such that MP gives different solutions than the MC formulation. It is shown that the set of Lagrange multipliers from the MC formulation plays an important role in separating the aforesaid regions. When the service charges are lower than the set of multipliers, the MC and MP formulations will give the same solution. When the service charges are higher than the multipliers, the solutions will be different. The same set of Lagrange multipliers also acts as thresholds for the service charges so that the SON operator would provide adequate service levels under the MP formulation. All these observations will be elaborated shortly. This paper is structured as follows: section 2 is the description of the problem assumptions and the employed analytic models, section 3 discusses the major results, section 4 shows the relation between this study and the pricing of services with GoS guarantees, a simple example illustrating the obtained results will be provided in section 5, and section 6 is the conclusion section that discusses the results obtained and possible future extensions.

2. PROBLEM DESCRIPTION

The SON network operates in a manner similar to a virtual network. The SON operator owns the SON gateways which are placed in strategic locations. To realize the SON network, the SON operator leases bandwidths with QoS guarantees from the underlying Autonomous Systems, (ASes) in the form of Service Level Agreements (SLAs). The leased bandwidths provide logical connections to the overlay network. Once the bandwidths are in place, the SON is realized and is ready to offer end-to-end QoS guarantees for the value-added services it provides (i.e. VoIP, Video On Demand services, etc). Users with access to the Internet can access the service gateways and use the value-added services. The SON connections are classified by the source and the destination gateways. Users pay the service charge based on the sources and destinations of their connections as well as their connection durations. To deploy a SON, a major challenge faced by the SON operator would be the optimal amount of bandwidth to be leased on each of the logical links. The allocated bandwidths should provide the operator with maximum economic benefit yet meet the user expectation regarding the connection acceptance ratio. This acceptance ratio is also known as the Grade of Service requirements (GoS) which are usually represented in the form of connection blocking probabilities. The problem confronting the SON operator is: given a set of SON gateways, decide the optimal bandwidths to be leased on the logical links when the traffic intensities and the gateway locations are known, and the service charges as well as GoS requirements are given.

The maximum profit formulation of this problem is given by the expression in (1).

$$\max_{N_s} \quad \sum_{ij} \lambda^{ij} w^{ij} (1 - B_{ij}) - \sum_{s} C_s(N_s)$$
$$B_{ij} \le \overline{L}^{ij} \qquad u_{ij}$$
$$N_s \ge 0 \qquad z_s \tag{1}$$

Where λ^{ij} is the poissonian connection arrival rate for the node pair *ij* (i.e. source node is *i*, destination node is *j*), w^{ij} is the expected reward generated by an admitted *ij* connection, both λ^{ij} and w^{ij} are assumed to be given parameters. The variable B_{ij} is the blocking probability for traffic pairs *ij*, which is a value returned by the routing layer. The GoS constraint for each OD pair *ij* is given by $\overline{L}^{ij}$ which specifies the maximum allowed connection blocking probability, it is assumed that the Erlang B equation is employed to quantify the blocking probabilities. The capacity of a link *s* is denoted by N_s and it is a decision variable of this problem. The function $C_s(.)$ is the cost function that quantifies the cost rate of allocating N_s units of capacities on link *s* and it is assumed to be a linear function of the variable N_s. The variables u_{ij} and z_s are the Lagrange multipliers with respect to the constraints. The minimum cost formulation of the same problem is given by (2).

$$\min_{N_s} \quad \sum_{s} C_s(N_s)$$
$$B_{ij} \le \overline{L}^{ij} \qquad v_{ij}$$
$$N_s \ge 0 \qquad z_s \tag{2}$$

It is important to mention that the blocking variable B_{ij} donates the theoretical blocking probabilities of the OD pair *ij* and these probabilities are dependent on the routing scheme employed and its performance model. For the same network configuration (same allocated capacities), different routing schemes may give different B_{ij} values. A routing scheme that utilizes the network resources efficiently may return lower B_{ij} values than a simple and primitive routing scheme even with the same set of allocated capacities.

3. THE MP AND MC FORMULATIONS

3.1 The Equivalence of MP and MC Formulations

From (1) and (2) the Maximum Profit (MP) and the Minimum Cost (MC) formulations are identical if the w^{ij}'s are all zero. Intuitively when the w^{ij}'s gradually increase, the cost would still dominate the objective until a threshold is hit. In other words, the MP and MC formulations should be equivalent when the w^{ij}'s are smaller than some thresholds. In this section we formally investigate this and show the conditions that the MP and MC formulations are equivalent. The results rely on the non-negative nature of a set of multipliers. The strictly convex property of the Erlang B equation is also employed to make the proof clearer. Only the results for the cases where the MC formulation has a unique solution are shown, and this is assumed throughout the whole article. The results for the multiple-solution cases will be included in an extended version of the paper.

Before showing the main result, we need to prove a lemma that establishes the uniqueness of the allocated capacity with respect to the cost, a set of real values v_{ij} (which mimic the Lagrange multipliers corresponding to the GoS constraints) and the link traffic intensity (i.e. offered traffic intensity) vector Λ.

Lemma 1:

For positive constants v_{ij}, c_s, and fixed Λ, there exists at most one N_s on the link *s* that satisfies the following equation:

$$c_s = \sum_{ij \in \theta_s} v_{ij} \left(-\frac{\partial B_{ij}}{\partial N_s}\right) \tag{3}$$

Where θ_s is a set containing the indexes of all Origin-Destination pairs that have link *s* in their routes, N_s is the capacity of link *s* and it is relaxed to be a continuous variable, c_s is the cost of allocating one unit of capacity on link *s*, v_{ij} are some positive real numbers, Λ is the vector of traffic intensities on the links as designated by some optimal routing rules.

Proof:

Assume the continuous extension of Erlang B equation suggested in [1] is being employed. B_{ij} can be written as (4), where R_{ij} is a set that contains all the links being used by the routes of traffic *ij*, the functions $f_1^{ij}(.)$ and $f_2^{ij}(.)$ are to characterize the blocking of traffic *ij* due to links other than link *s* in the routes, therefore they are both non-negative. The function $E_s(.)$ is the Erlang B equation that quantifies the blocking due to lack of resource on link *s*. The values λ_s are the elements of Λ which denotes the offered traffic intensity on link *s*. The idea for (4) is that the blocking probability of the traffic pair *ij* involves terms that are independent of link *s* and also terms that are dependent on link *s*. This idea is rather general and the expression (4) is valid for various routing schemes like the alternate routing scheme [3], [4], the load sharing routing scheme [5], and extended versions of the load sharing/alternate routing schemes [6]. Note that for expression (4) to be valid, the condition $s \in R_{ij}$ is needed. If this condition is not satisfied, (4) can be still employed but $f_2^{ij}(.)$ becomes zero.

$$B_{ij} = f_1^{ij}(\lambda_{\substack{s' \in R_{ij} \\ s' \ne s}}, N_{\substack{s' \in R_{ij} \\ s' \ne s}}) + f_2^{ij}(\lambda_{\substack{s' \in R_{ij} \\ s' \ne s}}, N_{\substack{s' \in R_{ij} \\ s' \ne s}}) E_s(\lambda_s, N_s) \tag{4}$$

By substituting (4) into (3) and by assuming link independence we have:

$$\frac{c_s}{\sum_{ij \in \theta_s} v_{ij} \times f_2^{ij}(\lambda_{\substack{s' \in R_{ij} \\ s' \ne s}}, N_{\substack{s' \in R_{ij} \\ s' \ne s}})} = \left(\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s}\right) \tag{5}$$

Since the functions $f_2^{ij}(.)$ are independent of N_s and λ_s, they can be regarded as a non-negative constant functions with values in the interval (0,1], note that 0 is not included since $ij \in \theta_s$. Denote $\left(\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s}\right)$ by $f_3(\lambda_s, N_s)$. It is known that the continuous Erlang B formula is a C^∞ function [2], which is strictly convex in the capacity [1]. Therefore for a fixed λ_s, the function $f_3(.)$ is strictly decreasing and continuous in N_s, and there is an one-to-one correspondence between the function's value and N_s for the fixed λ_s. Moreover it can be seen that $f_3(.)$ is positive (since when N_s is increased by delta, the function $-E_s(.)$ always increases if λ_s is fixed), so for positive constants v_{ij}, c_s, and fixed offered traffic intensity λ_s on the link, there is a unique N_s that satisfies equation (5). There are two special cases for expression (5), first if the LHS of (5) is larger than $\max_{N_s}(f_3(\lambda_s, N_s))$ then N_s does not exist, second if the cost c_s equals to zero (which is unlikely), then N_s equals infinity. **Q.E.D.**

Expression (3) together with the complementary slackness and the non-negativity requirement of the multipliers correspond to the 1st order necessary conditions of the MC formulation. The physical interpretation of lemma 1 is that when the link traffic intensities are fixed, then for each set of c_s and v_{ij}, there is at most one N_s value that solves (3) and thereby satisfying the 1st order necessary conditions. The result implies that we may write N_s as $N_s(c_s, v_{ij}, \Lambda)$. This lemma provides a tool for us to show the conditions such that the MC and MP formulations are equivalent. The proof of the following theorem relies on the non-negativity nature of the multipliers corresponding to the GoS.

Theorem 1 (relationship between the minimum cost and maximum profit formulation):

Consider a maximum profit formulation (1) and a minimum cost formulation (2) of the SON capacity allocation problem. If a set of Lagrange multipliers v_{ij}^* exists for (2), and if the connection rewards w^{ij} satisfy condition (6), then an optimal solution to (2) is also an optimal solution to (1). Further if there is a unique set of v_{ij}^* for (2), then (1) and (2) are equivalent.

$$0 \le w^{ij} \le \frac{v_{ij}^*}{\lambda^{ij}} \quad \forall ij \qquad (6)$$

Proof:

The stationary conditions for (1) and (2) are listed in (7) and (8) respectively, where c_s is a constant cost for allocation a unit bandwidth on link s. Note that (7) and (8) both represent n sets of equations where n equals to the number of links in the network. Again θ_s in (7) and (8) is a set that contains the indexes of all OD

$$c_s = \sum_{ij \in \theta_s} (\lambda^{ij} w^{ij} + u_{ij}^*)\left(\frac{-\partial B_{ij}}{\partial N_s^{p*}(\Lambda^{p*})}\right) \qquad (7)$$

$$c_s = \sum_{ij \in \theta_s} v_{ij}^* \left(\frac{-\partial B_{ij}}{\partial N_s^{c*}(\Lambda^{c*})}\right) \qquad (8)$$

pairs that consist of link s in their routes. Note that the capacities N_s^{c*} and N_s^{p*} are represented as functions of the link traffic intensities to explicitly indicate the dependence on the solution of the routing layer. Suppose v_{ij}^* are the Lagrange multipliers at optimality for (2). It is easy to see that if (6) is satisfied, letting $u_{ij}^* = v_{ij}^* - \lambda^{ij} w^{ij} \ge 0$ for all the OD pairs ij in all the n equations of (7) would make the values $\sum_{ij \in \theta_s} (\lambda^{ij} w^{ij} + u_{ij}^*)$ equal to $\sum_{ij \in \theta_s} v_{ij}^*$ for all the n equations. If (6) is not satisfied, this will not be always possible since u_{ij}^* must be non-negative for all ij. Assume a vector of link intensities Λ_1^{c*} is designated by some optimal routing rules, and suppose the tuple (Λ_1^{c*}, $N^{c*}(\Lambda_1^{c*})$) is a solution that satisfies (8) and also the complementary slackness conditions. Then it is obvious from lemma 1 that the solution ($\Lambda_1^{c*}, N^{c*}(\Lambda_1^{c*})$) is unique for the Λ_1^{c*}. Now if we substitute the solution along with the u_{ij}^* into (7), the solution should satisfy (7) if it satisfies (8), this is because that $(\lambda^{ij} w^{ij} + u_{ij}^*) = v_{ij}^*$ for all ij, and also because that the expression $\frac{-\partial B_{ij}}{\partial N_s^*(\Lambda^*)}$ depends only on Λ^* and $N^*(\Lambda^*)$. The complementary slackness conditions are also satisfied since the constraints of (1) and (2) are identical. In other words, when (6) is satisfied, a solution that satisfies the first order necessarily condition of the formulation (2) must also satisfy the 1st order necessarily condition of the formulation (1). With the same set of u_{ij}^* (i.e. $u_{ij}^* = v_{ij}^* - \lambda^{ij} w^{ij}$), the second order sufficient condition check is trivial, because the Hessian matrices of the corresponding Lagrangian duals are identical. If the solution ($\Lambda_1^{c*}, N^{c*}(\Lambda_1^{c*})$) satisfies the 2nd order sufficient conditions of formulation (2) then it must also satisfy the sufficient conditions of formulation (1). Therefore we have shown that the optimal solution of MC is also an optimal solution of MP. Further to this, if the set of v_{ij}^* is unique and condition (6) is satisfied, then formulations (1) and (2) are equivalent. This can be proved by contradiction. If there exists another set of multipliers x_{ij}^* such that $(\lambda^{ij} w^{ij} + x_{ij}^*) = y_{ij}^* \ne v_{ij}^*$ for the formulation (1), where $x_{ij}^* \ge 0$ and $v_{ij}^* \ge 0$. Suppose for any optimal solution ($\Lambda_2^{c*}, N^{c*}(\Lambda_2^{c*})$) that corresponds to x_{ij}^* and satisfies both the necessary and sufficient conditions of the formulation (1). It is obvious that the multipliers $y_{ij}^* = (\lambda^{ij} w^{ij} + x_{ij}^*)$ along with ($\Lambda_2^{c*}, N^{c*}(\Lambda_2^{c*})$) satisfy (8). Moreover this solution must also satisfy the second order necessarily conditions of (2) as long as it satisfies these conditions for (1), since formulations (1) and (2) have the identical first order and second order optimality conditions. That implies y_{ij}^* is also a set of multipliers for (2). This contradicts with the claim that (2) has a unique set of multipliers. Therefore if (6) holds and the set of multipliers v_{ij}^* is unique for (2), then (1) also have a unique set of multipliers u_{ij}^* such that $(\lambda^{ij} w^{ij} + u_{ij}^*) = v_{ij}^*$. And this shows the equivalence of the MP and the MC formulations. **Q.E.D.**

For the discussions in this article, the vector Λ^* is assumed to be unique with respect to a given set of multipliers. Yet this requirement is not needed for the above proof to be valid (but then we may have multiple optimal solutions even if the set of multipliers is unique). Assume there exists a unique set of multipliers for the MC formulation and the corresponding Λ^* is unique; the above theorem indicates that if the service charges are lower than some values (i.e. condition (6)), the MC and the MP formulations both give the same solution to the capacity allocation problem. In other words, if the service charges are lower than a set of thresholds, formulating the SON capacity allocation problem as either the MC or the MP problem does not matter as this is the region such that MC and MP are equivalent. A question one may ask is: what if the service charges are higher than the thresholds? The result for the question is presented in section 3.2.

3.2 The MP and MC Formulations as Differentiated by the Service Charges

If condition (9) holds for the service charges w^{ij}, then we shall show in this section that the MP formulation provides solution with *strictly* better performance than the MC formulation in terms of profit generation. Although some of the results obtained may look somewhat trivial, the formal investigation of the problem provides us with valuable insight into the relationship between the service charges and GoS guarantees which will be discussed in the section 4.

$$0 \le \frac{v_{ij}^*}{\lambda^{ij}} < w^{ij} \quad \forall ij \tag{9}$$

Note that though conditions (6) and (9) are somewhat complementary, they are in fact rather loose as there are "grey regions" which are not covered. These "grey regions" are the regions where $0 \le w^{ij} < \frac{v_{ij}^*}{\lambda^{ij}}$ for some ij, and $0 \le \frac{v_{ij}^*}{\lambda^{ij}} \le w^{ij}$ for the remaining ij. To tackle these regions, the study will involve the comparisons of the terms $R_s^1 = \sum_{ij\in\theta_s}(\lambda^{ij}w^{ij})$ and $R_s^2 = \sum_{ij\in\theta_s} v_{ij}^*$. Our results indicate that if there exists at least one R_s^1 that is larger than R_s^2 for some link s, then the MP formulation will give solution that generates more profit, otherwise the MP and MC formulations gives solutions that generates the same profit. The detailed analysis of these cases requires more careful treatment which needs significantly more space, these results will be included in a future paper as mentioned earlier. The result provided here is a simplified version based on the condition (9). Before we present the main results, again we first need a lemma. This lemma establishes the relation between the optimal capacities allocated and the magnitude of service charges.

Lemma 2:

Consider equation (10), where c_s is a positive constant. Assume that the link traffic intensities, Λ^*, are fixed. If v^1, v^2 are two vectors whose elements are indexed by ij (i.e. v^1=[v_{ij}^1]^T, v^2=[v_{ij}^2]^T), then if $v_{ij}^1 > v_{ij}^2 > 0$ for all ij, and $N_s^{v^1}$ and $N_s^{v^2}$ both exist then we have $N_s^{v^1} > N_s^{v^2}$, where $N_s^{v^1}$ and $N_s^{v^2}$ are the values of N_s^v in (10) that correspond to the vectors v^1 and v^2 respectively.

$$c_s = \sum_{ij\in\theta_s} v_{ij}\left(-\frac{\partial B_{ij}}{\partial N_s^v}\right) \tag{10}$$

Proof:

Substituting (4) into (10) we have:

$$c_s = \sum_{ij\in\theta_s} v_{ij} \times f_2^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\ne s}}, N_{\substack{s'\in R_{ij}\\ s'\ne s}})\left(\frac{-\partial E(\lambda_s, N_s)}{\partial N_s^v}\right) \tag{11}$$

From the proof of lemma 1, It is known that $\frac{-\partial E(\lambda_s, N_s^v)}{\partial N_s^v}$ is strictly decreasing and continuous for fixed λ_s, moreover it is also known that $f_2^{ij}(.)$ are non-negative constant functions with respect to N_s^y. So the expression $\sum_{ij\in\theta_s} v_{ij} \times f_2^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\ne s}}, N_{\substack{s'\in R_{ij}\\ s'\ne s}})\left(\frac{-\partial E(\lambda_s, N_s)}{\partial N_s^v}\right)$ is continuous and strictly decreasing in N_s^v. As a result for a constant c_s, the larger the function $\sum_{ij\in\theta_s} v_{ij} \times f_2^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\ne s}}, N_{\substack{s'\in R_{ij}\\ s'\ne s}})$ the smaller the expression $\left(\frac{-\partial E(\lambda_s, N_s)}{\partial N_s^v}\right)$ will be required to satisfy the equality condition of expression (11), this therefore requires a larger N_s^v value. It is easy to see that if $v_{ij}^1 > v_{ij}^2 > 0$ for all ij, the function $\sum_{ij\in\theta_s} v_{ij}^1 \times f_2^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\ne s}}, N_{\substack{s'\in R_{ij}\\ s'\ne s}})$ will be larger than $\sum_{ij\in\theta_s} v_{ij}^2 \times f_2^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\ne s}}, N_{\substack{s'\in R_{ij}\\ s'\ne s}})$, so if $N_s^{v^1}$ and $N_s^{v^2}$ both exist then $N_s^{v^1} > N_s^{v^2}$. **Q.E.D.**

With lemma 2, we can now continue with our main theorem of this section.

Theorem 2:

Assume the link traffic intensities at optimality, Λ^*, are the same for the MC and MP formulations. When condition (9) holds, the MP formulation delivers solution with *strictly* better performance than the MC formulation in terms of profit.

Proof:

Note that when condition (9) holds, u_{ij}^* in (7) are all zero at optimality, this is due to the theorem 3 in section 4 and the complementary slackness condition. Now expression (7) can be written as (12)

$$c_s = \sum_{ij\in\theta_s}(\lambda^{ij}w^{ij})\left(\frac{-\partial B_{ij}}{\partial N_s^{p^*}}\right) \tag{12}$$

From the expressions (8) and (12) and also lemma 2, it is easy to verify that when condition (9) holds the optimal capacity allocated by MP is strictly larger than that of MC on all links. To show the profit generated by the solution of MP formulation is strictly larger if condition (9) holds, we rewrite (9) as $0 \le \frac{v_{ij}^*}{\lambda^{ij}} + x_{ij}^* = w^{ij}$, where $x_{ij}^* > 0$ for all ij. We substitute this into (12) and rearrange the terms we have (13)

$$c_s = \sum_{ij\in\theta_s}(v_{ij}^*)\left(\frac{-\partial B_{ij}}{\partial N_s^{p^*}}\right) + \sum_{ij\in\theta_s}(\lambda^{ij}x_{ij}^*)\left(\frac{-\partial B_{ij}}{\partial N_s^{p^*}}\right) \tag{13}$$

we substitute the optimal solution of MC (i.e. N^{c^*}) into the RHS of (13) and we get (14),

$$\sum_{ij\in\theta_s}(v_{ij}^*)(\frac{-\partial B_{ij}}{\partial N_s^{c*}})+\sum_{ij\in\theta_s}(\lambda^{ij}x_{ij}^*)(\frac{-\partial B_{ij}}{\partial N_s^{c*}}) \quad (14)$$

expression (14) denotes the marginal profit from the link *s* at the allocated capacity of N^{c*}, and it can be further modified to (15)

$$c_s+\sum_{ij\in\theta_s}(\lambda^{ij}x_{ij}^*)(\frac{-\partial B_{ij}}{\partial N_s^{c*}}) \quad (15)$$

from the proof of lemma 1, we know that the 2nd term of the expression (15) is strictly positive, therefore expression (15) suggests that at the optimality of MC formulation (i.e. N^{c*}), marginal reward is larger than marginal cost on the links, which implies that additional profit can be generated if extra capacities are allocated. Therefore when (9) holds, the MP formulation provides strictly better performance in terms of profit generation. **Q.E.D.**

The result of theorem 2 is somewhat trivial, because after all, the MP formulation is to maximize the profit, but from the analysis in theorem 2, we can gain the addition insight that the solutions of the MC formulation offer worse profit than that of the MP formulation because the MC formulation merely allocates the minimum amount of capacities to satisfy the GoS requirements. Additional profit that could have been gained from the high reward connections is lost. The main conclusion from sub-sections 3.1 and 3.2 is that the MP formulation offers no worse performance than the MC formulation in all scenarios; when rewards are low and the costs are the major concern, the MP formulation minimizes the cost, when the rewards are high enough, the MP formulation switch to maximize the profit. Therefore it is sensible to design the SON network using the MP formulation. If the MP formulation is chosen, it can be further shown that the multipliers v_{ij}^* in (9) act as a set of threshold prices for the service so that the operator would be happy to deliver adequate level of GoS even if he is not constrained to do so. We shall elaborate this in the following section.

4. THE PRICE OF OFFERING GOS: AN OPERATOR'S PERSPECTIVE

In this section we assume that the SON operator regards the profit as the primary metric for the network performance. The following theorem formally shows that the multipliers v_{ij}^* act as a form of threshold for service charges in the MP formulation.

Theorem 3:

Assume that OD-pair blocking probabilities are strictly decreasing functions of allocated capacities. If (9) holds then the solution of MP formulation will deliver lower blockings than that of the MC formulation. Moreover all the GoS constraints in (1) will be non-binding at optimality, and the optimal solution for (1) will be the same as its unconstrained counterpart.

Proof:

if (9) holds then $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij}) > \sum_{ij\in\theta_s}v_{ij}^*$, by lemma 2, we have $N^{p*} > N^{c*}$, where N^{p*} is the capacity vector that satisfy (7). Since $N^{p*} > N^{c*}$, the GoS offered by N^{p*} must be better than that being offered by N^{c*}. It is known that all the GoS constraints are satisfied by the allocation N^{c*}, so it must be satisfied by N^{p*}. Since all the GoS requirements are satisfied at the optimal solution of (1) without the need of having the GoS constraints if (9) holds, so the solution of (1) remains the same even if all the GoS constraints are relaxed. **Q.E.D.**

Assume that the MC formulation has a unique set of multipliers v_{ij}^*.

For the sake of discussion we define a set of threshold values $\tilde{w}^{ij}$ in (16):

$$\tilde{w}^{ij}=\frac{v_{ij}^*}{\lambda^{ij}} \quad (16)$$

If the service charges w^{ij} are less than the $\tilde{w}^{ij}$ for all *ij*, then from theorem 1, the MP and MC formulations are equivalent, and the multipliers u_{ij}^* in (7) will be positive. From the complementary slackness conditions, the MP formulation will deliver a solution with tight GoS constraints. It can also be seen from theorem 3 that if the service charges w^{ij} are larger than $\tilde{w}^{ij}$, the MP formulation will deliver GoS service levels better than that are being required by the GoS constraints. The set of $\tilde{w}^{ij}$ acts as a threshold set for the GoS service levels. In other words, the values $\tilde{w}^{ij}$ can be interpreted as the minimum service prices that drive the network operator to offer the required level of GoS when they are not constrained to offer any GoS guarantee. This set of $\tilde{w}^{ij}$ provide interesting cost information in the context of providing GoS guarantees. The thresholds in (16) are derived from the multipliers v_{ij}^*, which are well known metrics that quantify the prices of the GoS constraints: the cost objective in (2) can be improved by v_{ij}^* units if the corresponding GoS constraint is relaxed by one unit. So, intuitively this is also the amount of reward the users of the OD pair *ij* should bring to the network so as to enjoy the GoS.

5. AN ILLUSTRATION EXAMPLE

Consider a simple three-node SON network as shown in figure 1. Assume that there are three Poisson streams of traffic in the network, and the offered traffic intensities for the three independent streams are λ_{AB}=10 per unit time, λ_{CB} =15 per unit time, λ_{AC} =20 per unit time. To make the discussion simple, all the traffic streams are routed through direct links (i.e. direct links AB, CB and AC). Without loss of generality the mean holding times of the traffic are assume to be identically distributed with unit mean. The costs for leasing one unit of bandwidths for one unit of time are 5 units, 6 units and 7 units respective for the links AB, CB, and AC, and the allocated capacities are assumed to be integral values. Assume that GoS requirements for all the streams are 0.1 (i..e.10% probability of blocking)), this value is deliberately made large so as to facilitate easy comparison. The multipliers v^*_{AB}, v^*_{CB} and v^*_{AC} are found to be 182, 267 and 372 respectively (which translates to service charges of 18.2, 17.8 and 18.6). To illustrate theorem 1, we deliberately set the service charges to be (10, 10, 10) for the streams AB,CB and AC. Condition (6) is satisfied under this set of service charges. Table 1 summarizes the results obtained. We can see that, under these service charges, the MP and MC formulations do give the same allocated capacities even though the objective values are different, which confirms the results in theorem 1.

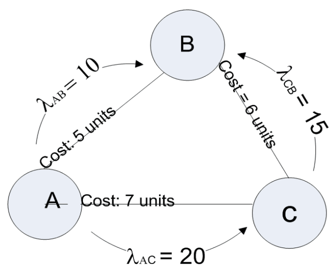

Figure 1. A simple SON network.

Table 1. Capacity allocation results for low service charge

Service charges (10,10,10)	**MP formulation**	**MC formulation**
GoS ($\lambda_{AB}, \lambda_{CB}, \lambda_{AC}$)	(0.084, 0.086, 0.085)	(0.084, 0.086, 0.085)
Allocated capacities on links (AB, CB ,AC)	(13, 18, 23)	(13, 18, 23)
Objective value	77.65	-334

To illustrate theorem 2, we set the service-charge vector to be (200, 200, 200). Table 2 summarizes the results obtained. The profit rate improvement of the MP over the MC is 596.24. The MP formulation delivers better performance over the MC formulation in this case. This is consistent with the results proved in theorem 2.

Table 2. Capacity allocation results for high service charge

Service charges (200,200,200)	**MP formulation**	**MC formulation**
GoS ($\lambda_{AB}, \lambda_{CB}, \lambda_{AC}$)	(0.0037,0.0029,0.0034)	(0.084,0.086,0.085)
Allocated capacities on links (AB, CB ,AC)	(19, 26, 32)	(13, 18, 23)
Expected Profit rate	8495.34	7899.10

Since the solutions are rounded up from the real-valued optimal solutions, the GoS constraints are not tight even for the MC formulation, but note that in table 2 when the service charges satisfy (9), the solution of the MP formulation indeed offers lower blocking probabilities than that of the MC formulation and this illustrates the results in theorem 3.

6. CONCLUSIONS AND FUTURE WORKS

We studied a class of Service Overlay Network (SON) capacity allocation problem with GoS constraints. By using the set of Lagrange multipliers in the MC formulation, we showed the condition such that the MP and MC formulations are equivalent. Moreover we also showed the condition such that the two formulations are different. It is particular interesting to note that when the service charges are low the MP formulation minimizes the cost/investment in realizing the SON networks by providing minimum level of resources to meet the GoS constraints. When the service charges are high enough, however, the MP formulation automatically takes advantage of that and switches to maximize the profit in realizing the SON network. This phenomenon suggests that MP formulation adapts well to the different ranges of parameters in the SON design problem and is attractive to be employed as the dimensioning formulation. By employing the MP formulation in the SON capacity allocation problem, we are able to quantify the prices of offering GoS in the SON. An interesting fact is that regardless that the MP formulation is being employed; the aforesaid GoS prices are related to the set of Lagrange multipliers of the MC formulation. Major efforts are being spent to refine the results obtained regarding the relationship of the MP and MC formulations when the MC formulation has multiple optimal solutions, the possible extension to the cases where heterogeneous blockings are present is also being studied. Moreover we are also investigating the extensions of the results to the more general multi-rate cases. Most of the new results will be included in a more comprehensive future paper.

7. REFERENCES

[1] A. Jagers and E. V. Doom "On the continued Erlang loss function", Operations Research Letters, Vol. 5, No. 1, pp. 43-46, 1986.

[2] J. S. Esteves, J. Craveirinha, and D. M. Cardoso, "Second order conditions on the overflow traffic from the Erlang-B system", Cadernos de Matemática, Universidade de Aveiro, CM06/I-20, 2006.

[3] A. Girard and B. Liau, "Dimensioning of adaptively routed networks", IEEE transactions on networking, Vol. 1, No. 4, pp. 460-468, 1993.

[4] A. Girard, "Revenue optimization of telecommunication networks", IEEE transactions on communications, Vol. 41, No. 4, pp. 583-591, 1993.

[5] F.P. Kelly, "Routing in circuit-switched networks: Optimization, Shadow Prices and Decentralization", Advanced in applied probability, Vol. 20, No. 1, pp. 112-144 , 1988.

[6] N. Lam, Z. Dziong and L.G. Mason, "Network capacity allocation in service overlay networks", in Proc. 20th International Teletraffic Congress (ITC), Ottawa, 2007, pp. 224-235.

[7] A. Girard, "Routing and dimensioning in circuit-switched networks", Addison-Wesley, 1990.

[8] D. P. Bertsekas, "Nonlinear Programming", 2nd edition, Athena Scientific, 1999.