MAXIMUM PROFIT VS MINIMUM COST IN SERVICE OVERLAY NETWORK DESIGN

ABSTRACT

We studied a class of Service Overlay Network (SON) capacity allocation problem with Grade of Service (GoS) constraints. The problem can be 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. We use a set of Lagrange multipliers to investigate the general conditions for the MP formulation to be equivalent to the MC formulation. The set of multipliers can also be shown to act as a set of thresholds for user 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. The 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 the network designers may consider.

KEYWORDS

Grade of Service guarantees, Network design, Optimization

1. INTRODUCTION

It is non-trivial to provide end-to-end Quality of Service guarantee in the Internet, as it consists of a large collection of independent Autonomous Systems (ASes). To ensure end-to-end QoS guarantees, one has to build a multi-lateral business relationship with all the independent ASes his data transits. A higher level mechanism on the top of the Internet known as Service Overlay Network (SON) is thus proposed to alleviate this problem [12]. 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 act as logical links that connect the SON gateways. Once all the logical links are in place, the SON is realized and the overlay network formed is under the administration of a single authority. Because the SON is administrated by a single operator, it is capable of providing end-to-end QoS guarantees for the value-added services provided by it (i.e. VoIP, Video conferencing, online games, etc). A user with access to the Internet can access the service gateway to use the value-added services, provided the hosts holding the contents are also connected to some SON service gateways. In a SON, the connections are classified by the origin and the destination (OD) gateways. Users pay the service charge based on the origins and destinations of their connections as well as their connection durations. It is quite common that in formulating SON network design problems, one may be confronted with the choice of choosing to maximize the profit or to minimize the cost, as the design criteria. An interesting question is thus raised: since the maximum profit (MP) and the minimum cost (MC) approaches optimize different objectives, how do their solutions compare and how does a network designer decide which approach to be employed. These questions could be non-trivial under the SON environment. It is because the SON operators enjoy great degree of freedom in deploying their own inter-domain routing schemes as the network is solely under their administrations. Owning to this flexibility, different network administrators are likely to employ different routing schemes that fit their business objectives. The routing scheme employed would influence the network design significantly. The conclusions drawn regarding SON design for a particular routing scheme may not be valid for another routing scheme. This paper addresses this objective choice problem faced by the operators for a class of Service Overlay Network (SON) capacity allocation problems. The result does not depend on a specific routing scheme as long as certain conditions are met. This paper is structured as follows: Section 2 is the description of the problem assumptions and formulations, Section 3 discusses the major results, Section 4 shows a simple example that illustrates the results, Section 5 is the conclusion section that discusses the conclusions obtained.

2. THE PROBLEM

2.1 Problem Formulations

To deploy a SON, a major challenge faced by the operator would be the optimal amount of bandwidths to be leased on the logical links. The allocated bandwidths should fit the economic objective of the operator yet meet the user expectation regarding the Grade of Service (GoS) levels. The Grade of Service (GoS) levels are usually quantified by connection blocking probabilities. By considering the SON as a connection oriented loss network [5], the SON design problem can be formulated as a constrained maximum profit problem (MP) as listed in (2.1a). The problem is assumed to be solved using the Lagrangian relaxation approach [8]. The corresponding first order optimality condition and the Hessian matrix for the relaxed problem are listed in (2.1b) and (2.1c) respectively. We shall denote formulation (2.1a) as MP in the remainder of this article.

$$\begin{aligned} \max_{N_s} \quad & \sum_{ij} \lambda^{ij} w^{ij} (1-B^{ij}) - \sum_s C_s(N_s) \\ & B^{ij} \le \bar{L}^{ij} \qquad u_{ij} \\ & N_s \ge 0 \qquad z_s \end{aligned} \qquad (2.1a)$$

$$c_s = \sum_{ij} (\lambda^{ij} w^{ij} + u^*_{ij})(\frac{-\partial B^{ij}}{\partial N^*_s}) \quad \forall \text{s, s.t. } N^*_s > 0 \qquad (2.1b)$$

$$H = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nn} \end{bmatrix} \quad h_{nm} = \sum_{ij} (w^{ij} \lambda^{ij} + u^*_{ij}) \frac{\partial^2 B^{ij}}{\partial N_n \partial N_m} \; , \; h_{nn} = \sum_{ij} (w^{ij} \lambda^{ij} + u^*_{ij}) \frac{\partial^2 B^{ij}}{\partial N_n^2} \qquad (2.1c)$$

In the formulation (2.1a), λ^{ij} is the given poissonian connection arrival rate demanding the connection of the node pair *(i,j)* (i.e. origin gateway is *i*, destination gateway is *j*), w^{ij} is the given expected reward generated by an admitted *(i,j)* connection. The symbol B^{ij} denotes the analytical end-to-end blocking probability for connections requesting connecting the node pair *(i,j)* due to lack of available resource. It is an end-to-end blocking probability dependent on the (optimal) routing scheme employed, which usually has a rather complex functional form [4]. The GoS constraint for each OD pair *(i,j)* is given by $\bar{L}^{ij}$, which specifies the maximum allowed end-to-end blocking probability. 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* (based on some SLA) 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 (MC) of the same problem, the first order optimality condition and the Hessian matrix of the relaxed problem are given by (2.2a), (2.2b) and (2.2c) respectively. The only new notation introduced in (2.2a) is v_{ij}. It is the Lagrange multiplier corresponds to the GoS constraint. We shall denote formulation (2.2a) as MC in the rest of the article. It is assumed that both MP and MC formulations employ the same (optimal) routing scheme in the routing layer.

$$\begin{aligned} \min_{N_s} \quad & \sum_s C_s(N_s) \\ & B^{ij} \le \bar{L}^{ij} \qquad v_{ij} \\ & N_s \ge 0 \qquad z_s \end{aligned} \qquad (2.2a)$$

$$c_s = \sum_{ij} v^*_{ij} (\frac{-\partial B^{ij}}{\partial N^*_s}) \qquad \forall \text{s, s.t. } N^*_s > 0 \qquad (2.2b)$$

$$H = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ \vdots & \ddots & \vdots \\ h_{n1} & \cdots & h_{nn} \end{bmatrix} \quad \text{where } h_{nm} = \sum_{ij} v^*_{ij} \frac{\partial^2 B^{ij}}{\partial N_n \partial N_m} \; , \; h_{nn} = \sum_{ij} v^*_{ij} \frac{\partial^2 B^{ij}}{\partial N_n^2} \qquad (2.2c)$$

To analyze the two different formulations, a strict forward way is investigate the corresponding optimality conditions since they provide the richest information regarding the optimal solutions. From (2.1b) and (2.2b), one can see that the first order necessary conditions of the two formulations are equivalent if the expressions v^*_{ij} and $\lambda^{ij} w^{ij} + u^*_{ij}$ are equal for all the OD pairs (*i,j*), provided the routing dependent functions B^{ij} are the same. In this case the second order sufficient conditions of the MP and MC will also be the same, as the Hessian matrices (2.1c) and (2.2c) are identical then. Therefore, regardless of their different objectives, MP and MC *could* deliver the same solutions if one can set the multipliers u_{ij}^* appropriately, subject to the dual feasibility conditions.

When the solutions of MP and MC coincide, it indicates the cost function dominates the reward function so that the incorporation of additional reward information does not differentiate the solutions of MP from the solutions of MC. We shall show that under some specific range of reward parameters w^{ij}, the costs dominate the design. While in other range of the w^{ij} parameters, the reward parameters are large enough to influence the network design, so that the solutions by MP and MC are different.

2.2 The End-to-End Blocking Probabilities

Before any analysis can be made, a proper analytical form of the blocking function, B^{ij}, is needed. The actual functional form of B^{ij} varies with routing schemes [4]. We take a different perspective and derive it by using the insight that the blocking function can be approximated based on the link traffic intensities (at equilibrium) and the capacities [11]. Techniques from the reliability theory [9] were used to devise a general and functional form for B^{ij}, regardless of the actual routing scheme employed. The B^{ij} function obtained below is based on the reduced load approximation model [11], which assumed statistical link independence and Poisson link arrival rates.

We consider the *collection* of network paths, that connect a particular origin node *i* with a particular destination node *j*, as a complete system. The task of this system is to serve the connections between the node pair (*i,j*). Assume the network links are independent of one another. The links in the *collection* of paths are the independent components of the system. Denote these links by *s* and let R_{ij} be a set that contains all these links. Define an indicator variable y_s for the link *s*, whereas y_s equals to *zero* if link *s* has enough resource to admit at least one connection, and y_s equals to *one* if link *s* does not have resource to serve any connection. The expected value of y_s is therefore the blocking probability of link *s*. According to the reliability theory [9], a Boolean function $\phi(Y)$ that indicates whether the system has the available resource for new (*i,j*) connections can be defined by taking Y=[y_s] as the input. The complement of it, $\bar{\phi}(Y) = 1 - \phi(Y)$ is another Boolean function that indicates whether the system has ran out of resource for new (*i,j*) connections. Thus the expected value of $\bar{\phi}(Y)$ is the end-to-end blocking probability for connection pair (*i,j*). Since y_s are independent zero-one random variables and $\bar{\phi}(Y)$ is a Boolean function, we can perform the Shannon decomposition on the function $\bar{\phi}(Y)$. By using link *s* as the pivot we have (2.3).

$$\bar{\phi}(Y) = y_s\bar{\phi}(1_s, Y) + (1 - y_s)\bar{\phi}(0_s, Y) = \bar{\phi}(0_s, Y) + [\bar{\phi}(1_s, Y) - \bar{\phi}(0_s, Y)]y_s \tag{2.3}$$

Where $(0_s, Y)$ and $(1_s, Y)$ are the status vectors that differs only in the s^{th} link. The functions $\bar{\phi}(0_s, Y)$ and $\bar{\phi}(1_s, Y)$ indicates that whether the system has been blocked given that the link *s* is in admissible status/has been blocked. By the definition of y_s, the expectation E[y_s] is the blocking probabilities of link *s*. Assume the links are independent and link arrival rates are Poisson, we have expression (2.4). The expectations E[y_s] and E[$y_{s'\neq s}$] are replaced by the Erlang-B Loss formula E_s(.) and the vector $E_{s'\neq s}$ respectively in (2.4). The vector $E_{s'\neq s}$ denotes the collection of Erlang-B loss functions for all the links *s'* such that $s'\neq s$. The continuous extension of Erlang-B formula suggested in [1] is being used throughout this article which is given by (2.5)

$$E[\bar{\phi}(Y)] = f_{1,s}^{ij}(E_{s'\neq s}) + f_{2,s}^{ij}(E_{s'\neq s})E_s(.) \tag{2.4}$$

It should be clear now that B^{ij}=E[$\bar{\phi}(Y)$] is a reduced-load approximation of the end-to-end blocking probability, as link independence and Poisson link arrival rates are assumed. Note that $f_{2,s}^{ij} \geq 0$, if the end-to-end blocking probability B^{ij} is strictly decreasing in the presence of additional available link*. This is a

* $f_{2,s}^{ij}$ is the Birnbaum's importance measure of link *s* in the context of reliability theory.

monotonic property we imposed on the routing schemes and it is assumed throughout the article. We also impose the assumption that the routing objective function being uni-modular with respect to the capacities.

$$E_s(\lambda_s, N_s) = \{\lambda_s \int_0^{+\infty} e^{-\lambda_s z}(1+z)^{N_s} dz\}^{-1} \tag{2.5}$$

Since the value of the Erlang-B formula can be uniquely determined by the link capacity and the link traffic arrival rate [4], therefore (2.4) is rewritten to (2.6) to explicitly state the dependence of B^{ij}'s on the (equilibrium) link traffic intensities and link capacities.

$$B^{ij} = f_{1,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}}) + f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}}) E_s(\lambda_s, N_s) \tag{2.6}$$

Expression (2.6) is valid for all the link s. So we can represent B^{ij} in the form of $f_{1,s}^{ij} + f_{2,s}^{ij} E_s(\lambda_s,N_s)$ for every link s, where $f_{1,s}^{ij}$ and $f_{2,s}^{ij}$ are independent of the link s. In a similar spirit, and by applying the Shannon decomposition again on expression (2.6), B^{ij} can be further decomposed in terms of the blocking probabilities of link s and link s', as shown in (2.7).

$$B^{ij} = f_{1,s}^{ij}(0_{s'}) + f_{2,s}^{ij}(0_{s'})E_s(.) + (f_{1,s}^{ij}(1_{s'}) - f_{1,s}^{ij}(0_{s'}))E_{s'}(.) + (f_{2,s}^{ij}(1_{s'}) - f_{2,s}^{ij}(0_{s'}))E_s(.)E_{s'}(.) \tag{2.7}$$

The symbols $0_{s'}$ and $1_{s'}$ in (2.7) indicate the statuses of the link s'. For an instance, $f_{1,s}^{ij}(1_s)$ indicates the function $f_{1,s}^{ij}(.)$ given that link s' has been blocked. The functions $f_1^{ij}(.)$ and $f_2^{ij}(.)$ in (2.7) are independent of both link s and s'. By using expressions (2.6), the optimality conditions (2.1b) and (2.2b) are rewritten to (2.8) and (2.9).

$$[\sum_{ij}(\lambda^{ij}w^{ij} + u_{ij}^*) \times f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}})](\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s^*}) \quad \forall s, \text{ s.t. } N_s^* > 0 \tag{2.8}$$

$$[\sum_{ij} v_{ij}^* \times f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}})](\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s^*}) \quad \forall s, \text{ s.t. } N_s^* > 0 \tag{2.9}$$

Assume the continuous Erlang-B formula in [1] is employed. Then the second derivatives exist as the Erlang-B formula is a C^∞ function [2]. By using (2.6) and (2.7), the elements of the Hessian matrices of (2.1c) and (2.2c) are rewritten to (2.10) and (2.11) respectively.

$$h_{nm} = \sum_{ij\in\theta_n, ij\in\theta_m} (w^{ij}\lambda^{ij} + u_{ij}^*)(f_{2,n}^{ij}(1_m) - f_{2,n}^{ij}(0_m)) \frac{\partial E_n(.)}{\partial N_n} \frac{\partial E_m(.)}{\partial N_m} \quad , \; h_{nn} = \sum_{ij\in\theta_n} (w^{ij}\lambda^{ij} + u_{ij}^*) f_{2,n}^{ij}(.) \frac{\partial^2 E_n(.)}{\partial N_n^2} \tag{2.10}$$

$$h_{nm} = \sum_{ij\in\theta_n, ij\in\theta_m} v_{ij}^*(f_{2,n}^{ij}(1_m) - f_{2,n}^{ij}(0_m)) \frac{\partial E_n(.)}{\partial N_n} \frac{\partial E_m(.)}{\partial N_m} \quad , \; h_{nn} = \sum_{ij\in\theta_n} v_{ij}^* f_{2,n}^{ij}(.) \frac{\partial^2 E_n(.)}{\partial N_n^2} \tag{2.11}$$

3. THE EQUIVALENCE OF THE MP AND THE MC FORMULATIONS

Note that from the new optimal conditions (2.8) and (2.9), if the sums $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij} + u_{ij}^*) \times f_{2,s}^{ij}(.)$ and $\sum_{ij\in\theta_s} v_{ij}^* \times f_{2,s}^{ij}(.)$ can be made identical for all the links at the optimal solution of MC, the first order optimality conditions of MP and MC will be identical. The second order optimality conditions will then also be identical. To see this, we compare the corresponding terms of the Hessian matrices in (2.1c) and (2.2c) by using (2.10) and (2.11). Recall that $f_{2,s}^{ij}(0_{s'})$ and $f_{2,s}^{ij}(1_{s'})$ are the function $f_{2,s}^{ij}$ given that link s' is admissible/has been blocked. So $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij} + u_{ij}^*) \times f_{2,s}^{ij}(.) = \sum_{ij\in\theta_s} v_{ij}^* \times f_{2,s}^{ij}(.)$ implies $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij} + u_{ij}^*) \times f_{2,s}^{ij}(y_{s'}) = \sum_{ij\in\theta_s} v_{ij}^* \times f_{2,s}^{ij}(y_{s'})$ where $y_{s'}$ can be either 0 or 1. It is easy to see the Hessian matrices of the two formulations are identical. Hence the second order optimality conditions are also identical. Therefore the problem now amounts to showing the existence of a set of strictly positive (dual feasible) multipliers u_{ij}^*, such that the two aforesaid sums can be made equal. We shall show the general conditions for the multipliers to exist using the Farkas's lemma [8], instead of the regularity condition. No assumption regarding the number of OD pairs and network links is made. Without loss of generality, let's assume the OD pairs are indexed from *1* to $|ij|$, where $|ij|$ is number of distinct OD pairs. Let's also assume the links are indexed from *1* to $|s|$, where $|s|$ is the total number of links in the network. Define a matrix ***A*** and a vector ***B*** as in (3.1). Then the non-negative matrix ***A*** that has a dimension of $|s|\times|ij|$. The function $f_{2,s}^{ij=k}(.)$ in the matrix ***A*** indicates that it is contained in the blocking function of the k^{th} OD pair and it is the f_2^{ij} function of the link *s*. The function takes the optimal solution of the MC formulation as input. The vector ***B*** has a dimension of $|s|\times 1$. The problem of finding a set of non-negative u_{ij}^{*} is equivalent to finding a strictly positive vector ***U**** such that ***AU*=B***. Obviously a necessary (but not sufficient) condition is that vector ***B*** must be strictly positive, as all the elements of ***A*** are non-negative.

$$A=\begin{bmatrix} f_{2,1}^{ij=1}(.) & \cdots & f_{2,1}^{ij=|ij|}(.) \\ \vdots & \ddots & \vdots \\ f_{2,|s|}^{ij=1}(.) & \cdots & f_{2,|s|}^{ij=|ij|}(.) \end{bmatrix} \quad B=\begin{bmatrix} \sum_{ij\in\theta_1}(v_{ij}^{*}-\lambda^{ij}w^{ij})\times f_{2,1}^{ij}(.) \\ \vdots \\ \sum_{ij\in\theta_s}(v_{ij}^{*}-\lambda^{ij}w^{ij})\times f_{2,|s|}^{ij}(.) \end{bmatrix} \tag{3.1}$$

$$A^{T}p=\left[\sum_{s}p_s f_{2,s}^{ij=1}(.)\ ,\cdots,\ \sum_{s}p_s f_{2,s}^{ij=|ij|}(.)\right]^{T}=[\eta^{1},\ldots,\eta^{|ij|}]^{T}\ \geq 0^{T} \tag{3.2}$$

By the Farkas's lemma, there exists a strictly positive ***U****, if and only if we can show that for every vector ***P*** such that $\boldsymbol{A^{T}P\geq 0}$, the inner product $\boldsymbol{B^{T}P}$ is non-negative. Which is shown in expression (3.3). Note that

$$<B,p>=\sum_{s}p_s\sum_{ij}(v_{ij}^{*}-\lambda^{ij}w^{ij})\times f_{2,s}^{ij}(.)=\sum_{s}p_s\sum_{ij}d^{ij}\times f_{2,s}^{ij}(.)=\sum_{ij}d^{ij}\sum_{s}p_s\times f_{2,s}^{ij}(.)=\sum_{ij}d^{ij}\eta^{ij} \tag{3.3}$$

$d^{ij}=(v_{ij}{}^{*}-\lambda^{ij}w^{ij})$ are not restricted in sign. If (3.3) is always non-negative, then there exists a set of Lagrange multipliers ***U**** such that the MP formulation can be made equivalent to the MC formulation. Looking back at (3.1) and (3.2), it is possible to set some η^{ij} arbitrarily large by properly picking value(s) for p_s for some instances of ***A***. Since d^{ij} are not restricted in sign, the inner product of (3.3) can not be always non-negative unless some further constraints are imposed on the routing matrix ***A***. The only case that (3.3) is always positive, regardless of the ***A*** matrix, occurs when all the d^{ij} are non-negative. This implies condition (3.4). This is also the result one can obtain by using the regularity assumption.

$$0\leq w^{ij}\lambda^{ij}\leq v_{ij}^{*} \qquad \forall ij \tag{3.4}$$

Condition (3.4) is the sufficient condition such that the MP and MC formulations can be made equivalent, regardless of the routing scheme employed. So the following results are valid for all routing schemes that are monotonic and uni-modular. We first need to establish the uniqueness of the allocated link capacity with respect to the link cost, a set of real values v_{ij}, and the link traffic intensity vector Λ_s. Define θ_s to be a set which contains indexes of all the OD pairs that utilize link *s* in at least one of their paths. Let N_s be the capacity of a link *s*, which is assumed to be continuous. Let c_s be the cost of allocating one unit of capacity on link *s*. Let $E_s(\lambda_s,N_s)$ denote the Erlang-B formula with offered traffic intensity λ_s and link capacity of N_s.

Lemma 1

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

$$c_s=\sum_{ij\in\theta_s}v_{ij}\times f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}},N_{\substack{s'\in R_{ij}\\ s'\neq s}})(\frac{-\partial E_s(\lambda_s,N_s)}{\partial N_s}) \tag{3.5}$$

Proof:

Assume the continuous extension of Erlang-B formula suggested in [1] is being employed. Note that the function $f_2^{ij}=\mathrm{E}[\bar{\phi}(1_s,Y)-\bar{\phi}(0_s,Y)]$ is the same as the one defined earlier and it is strictly positive (since we imposed the condition $ij\in\theta_s$). The values λ_s are the elements of Λ_s. We rearrange the terms and get (3.6):

$$\frac{c_s}{\sum_{ij\in\theta_s} v_{ij}\times f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}})} = (\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s}) \tag{3.6}$$

Because the functions $f_{2,s}^{ij}$ is independent of N_s, it can be regarded as a positive constant functions with respect to N_s. Denote $(\frac{-\partial E_s(\lambda_s,N_s)}{\partial N_s})$ 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 (because when N_s is increased by delta, the blocking 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 (3.6). There are two special cases for expression (3.6), first if the LHS of (3.6) 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, then N_s equals infinity. □

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 (3.5) and thereby satisfying the 1st order necessary conditions. Lemma 1 provides a tool for us to show the conditions such that MC and MP are equivalent. The following theorem will establish the relationship between MC and MP under some specified conditions.

Theorem 1

Consider a MP formulation (2.1a) and a MC formulation (2.2a) of the SON capacity allocation problem. If a set of Lagrange multipliers v_{ij}^* exists for the MC formulation, and if the condition (3.4) is satisfied, then an optimal solution of the MC formulation is also an optimal solution of the MP formulation. Further if the set of v_{ij}^* is unique, then MP and MC formulations are equivalent.

Proof:

The stationary conditions for MP and MC are (3.7) and (3.8) respectively. All the symbols have been defined earlier. Note that (3.7) and (3.8) both represent n sets of equations where n equals to the number of links in the network.

$$c_s = [\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij} + u_{ij})\times f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}})](\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s}) \tag{3.7}$$

$$c_s = [\sum_{ij\in\theta_s} v_{ij}\times f_{2,s}^{ij}(\lambda_{\substack{s'\in R_{ij}\\ s'\neq s}}, N_{\substack{s'\in R_{ij}\\ s'\neq s}})](\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s}) \tag{3.8}$$

Suppose v_{ij}^* are the Lagrange multipliers at optimality for MC. If (3.4) is satisfied, then it is possible to find a vector of non-negative $\boldsymbol{U^*}$=[u_{ij}^*]such that $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij} + u_{ij}^*)\times f_{2,s}^{ij}(\lambda^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}}, N^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}})$ equal to $\sum_{ij\in\theta_s} v_{ij}^*\times f_{2,s}^{ij}(\lambda^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}}, N^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}})$ for all the n pairs of equations in (3.7) and (3.8). Assume a vector of link intensities $\Lambda_s^{c^*}$ is designated by some optimal routing rules, and suppose the tuple ($\Lambda_s^{c^*}, N^{c^*}(\Lambda_s^{c^*})$), where $N^{c^*}(\Lambda_s^{c^*}) = [N_s^{c^*}(\Lambda_s^{c^*})]$, is a solution that satisfies (3.8) and also the complementary slackness conditions. Then it is obvious from Lemma 1 that the solution ($\Lambda_s^{c^*}, N^{c^*}(\Lambda_s^{c^*})$) is unique with respect to $\Lambda_s^{c^*}$. Now if we substitute the solution ($\Lambda_s^{c^*}, N^{c^*}(\Lambda_s^{c^*})$), along with u_{ij}^* into (3.7), then (3.7) should be satisfied if (3.8) is satisfied. This is because that $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij} + u_{ij}^*)\times f_{2,s}^{ij}(\lambda^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}}, N^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}}) = \sum_{ij\in\theta_s} v_{ij}^*\times f_{2,s}^{ij}(\lambda^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}}, N^{c^*}_{\substack{s'\in R_{ij}\\ s'\neq s}})$ for all the n equations, and also because that the expression $\frac{-\partial E_s(\lambda_s, N_s)}{\partial N_s^*(\Lambda_s^*)}$ depends only on Λ_s^* and $N_s^*(\Lambda_s^*)$. The complementary slackness conditions are also satisfied since the constraints of MP and MC are identical. In other words, when (3.4) is satisfied, a solution that satisfies the first order necessary condition of MC must also be able to satisfy the first order necessary condition of MP. With the same set of u_{ij}^*, the second order sufficient condition check is trivial, because the Hessian matrices of the corresponding Lagrangian duals are identical (see (2.10) and (2.11)). 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 (3.4) is satisfied, then formulations MP and MC are equivalent. This can be proved by contradiction. If there exists another set of multipliers ***X****=[x^*_{ij}] such that $\sum_{ij\in\theta_s}(\lambda^{ij}w^{ij}+x^*_{ij})\times f^{ij}_{2,s}(.)=\sum_{ij\in\theta_s} y^*_{ij}\times f^{ij}_{2,s}(.)\neq\sum_{ij\in\theta_s} v^*_{ij}\times f^{ij}_{2,s}(.)$ for MP, where $x^*_{ij}\ge 0$ and $y^*_{ij}\ge 0$. Then the corresponding optimal solution, $(\Lambda^{p*}_s, N^{p*}_s(\Lambda^{p*}_s))$, satisfies both the necessary and sufficient conditions of the formulation (2.1a). It is obvious that the multipliers $y^*_{ij}=(\lambda^{ij}w^{ij}+x^*_{ij})$ along with $(\Lambda^{p*}_s, N^{p*}(\Lambda^{p*}_s))$ satisfy (3.8). The complementary slackness conditions are also satisfied because (2.1a) and (2.2a) have the same constraints. So the first order optimality condition of MC will be satisfied by this solution. Moreover this solution must also satisfy the second order necessary conditions of MC as long as it satisfies these conditions for MP. That implies ***Y****=[y^*_{ij}] is also a set of multipliers for MC. But this contradicts with the claim that MC has a unique set of multipliers. Therefore it is impossible for the set of x^*_{ij} to exist. As a result MP also has a unique set of multipliers (i.e. ***U****=[u^*_{ij}]). This shows that both MP and MC give the same optimal solution, and are therefore equivalent. □

Assume there exists a unique set of multipliers for the MC formulation and the corresponding Λ^*_s is unique; the above theorem indicates that if the service charges are lower than some values (i.e. (3.4) is satisfied), MC and MP give the same solution. For the other regions, MC and MP will be (strictly) different. It can be shown that the MP formulation offers a solution with strictly more profit and strictly better GoS guarantees in these regions. Owing to the space limitation, these results are discussed in an extended version of the current article, the monotone property of the routing scheme will be critical in showing the results.

4. AN ILLUSRATING EXAMPLE

Consider a simple SON network as shown in figure 1. Assume the offered traffic intensities for the three independent Poisson streams are λ_{AB}=10 units per unit time, λ_{CB} =15 units per unit time, λ_{AC} =20 units per unit time. To make the discussion simple, all the traffic streams are routed through direct links. Without loss of generality the mean holding times of the traffic are assume to be identically distributed with unit mean. The costs of leasing one unit of bandwidth for one unit of time are 5 units, 6 units and 7 units respective for links AB, CB, and AC, 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). We set the service charges to be (10, 10, 10) for the streams AB,CB and AC. Condition (3.4) is satisfied under this set of service charges, as all the d^{ij} are positive. Table 1 summarizes the results obtained. It is noted that under these service charges, the MP and MC formulations exactly give the same allocated capacities even though the objective values are different, which illustrates the result in theorem 1.

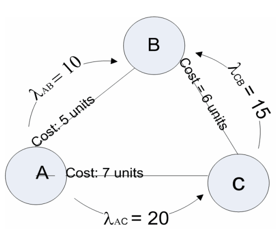

Figure 1. A simple SON network.

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)
Cost	334	334
Objective value	-77.65	334
Expected Profit rate	77.65	77.65

If, on the other hand, we set the service-charge vector such that it is profitable to offer the GoS (strict condition for "profitable" is not shown due to space limitation). The solution of MP formulation offers strictly lower blocking probabilities than that of the MC formulation and the profit of MP design surpasses that of the MC design.

5. CONCLUSIONS AND FUTURE WORKS

We studied a class of Service Overlay Network (SON) capacity allocation problem with GoS constraints. We showed the condition such that the MP and MC formulations are equivalent. Intuitively one can view the region that the MP and MC are equivalent to be the region that is not profitable to offer the required GoS guarantees. In this region the MP formulation minimizes the cost of offering GoS just like the MC formulation. In the profitable regions, however, MP takes advantage of that and allocates more resource whenever necessary to avoid potential profit losses (results not shown due to space limitation). This is an attractive feature because the MP approach automatically "decides" whether to minimize the cost or to maximize the profit in an optimal sense. The MC approach, on the other hand, does not have this capability. In cases that GoS requirements are based on some non-economic considerations (for example GoS can be decided by external entities like the regulating authorities), the MP approach will provide solutions with better values than that of the MC approach as the MP approach will be sticking blindly to the potentially arbitrary GoS requirements. The major computation efforts for both formulations will be on solving the optimal routing sub-problem and on calculating the gradients of B^{ij}. The computational costs for both models are therefore similar.

REFERENCES

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

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

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

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

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

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

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

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

[9] Wallace R. Blischke, D.N. Prabhakar Murthy, 2000. *Reliability Modeling, Prediction, and Optimization.* John Wiley & Sons, USA.

[10] Sheldon M. Ross, 2000. *Introduction to Probability Models*, 7th edition, Academic Press, USA.

[11] W. Whitt, 1985. Blocking when Service is Required from Several Facilities Simultaneously, *AT&T Technical Journal*, vol. 64, No. 8, pp. 1807-1856.

[12] Z. Duan, Z. L. Zhang, and Y. T. Hou, 2003. Service Overlay Networks: SLAs, QoS, and bandwidth provisioning, *IEEE/ACM Transactions on networking*, Vol.11, No. 6, pp. 870-883.