Bandwidth Allocation for a Revenue-Aware Network Utility Maximization

David Mayer and Javier A. Barria, *Member, IEEE*

***Abstract*— In this letter, we consider the Network Utility Maximization (NUM) problem in which a loss-related revenue indicator is set up by an operator as a constraint. We propose an algorithm which artificially modifies capacity constraints of a standard NUM algorithm. This results in a price-based mechanism that drives users into an operating point which maximizes total user utility for a given revenue indicator constraint.**

***Index Terms*— Bandwidth allocation, network utility maximization, revenue optimization, loss network.**

I. Introduction

BANDWIDTH allocation has been extensively studied with various objectives in view, especially objectives concerning the users (e.g. utility maximization and end-to-end delay minimization) and objectives of the operator's interest (e.g. revenue maximization and cost minimization). The question of combining the users' and operators' concerns remains an open issue. The motivation for this work stems from the fact that selfish user behavior comes at high cost for the operator and is conflicting with Traffic Engineering efforts [1].

In this letter, we attempt to solve the problem of Network Utility Maximization (NUM) with a loss-related constraint set up by the operator. Traditional NUM algorithms [2][3] result in fully utilized links. Assuming that users only pay for successfully delivered traffic, the loss associated with fully utilized links reduces operator's revenue. Hence we choose as a loss-related metric an operator's revenue indicator and solve the problem of NUM with this revenue indicator used as a constraint. To our knowledge, there is not a published work on trade-offs between utility and revenue in loss networks. However, several papers exist on optimization involving both revenue and some aspect of utility, typically fairness (e.g. [4]).

Our approach draws from the work of Mitra [5] on loss networks and uses a price-based mechanism to control user demand [2].

II. Model and Problem Formulation

Consider a set $\mathcal{L}$ of L links and a set $\mathcal{F}$ of F flows. Associated with each link $l \in \mathcal{L}$ there is a loss function $L_l(x_l)$, where x_l is the flow on link l. Link l capacity is denoted by C_l. Flow configuration is given and represented by routing matrix $\mathbf{A} = (a_{i,l}; i \in \mathcal{F}, l \in \mathcal{L})$, where $a_{i,l} = 1$ if flow i uses link l and $a_{i,l} = 0$ otherwise. Each flow i uses one path denoted by $R_i = \{l \in \mathcal{L} | \mathbf{A}_{i,l} = 1\}$. Flow $i \in \mathcal{F}$ is associated with a user i and a utility function $U_i(f_i)$, where f_i is the bandwidth demand of user i. Let $\mathbf{f} = (f_i, i \in \mathcal{F})$. We assume that utility functions are concave increasing.

Our goal is to optimize the capacity constraints provided to the standard NUM algorithm ([2][3]) in order to maximise total user utility U achieved by the NUM algorithm subject to the operator's revenue indicator π being equal to or greater than a positive constant π_{min}. The optimization problem of the operator is

π-NUM:

$$\max_{\mathbf{k} \geq 0} U(\mathbf{k}) \quad (1)$$
$$\text{s.t.} \quad \pi_{min} - \pi(\mathbf{k}) \leq 0,$$

where $\mathbf{k} = (k_1, \dots, k_L)$ is the control variable and $U(\mathbf{k})$ is the maximum of the following utility maximising problem.

NUM($\mathbf{k}$):

$$U(\mathbf{k}) = \max_{\mathbf{f} \geq 0} \sum_{i \in \mathcal{F}} U_i(f_i) \quad (2)$$
$$\text{s.t.} \quad \sum_{i \in \mathcal{F}} (f_i a_{i,l} - k_l C_l) \leq 0, \quad \forall l \in \mathcal{L}, k_l \in \mathbf{k}.$$

Denote $\mathbf{f}^*(\mathbf{k})$ the maximiser of NUM($\mathbf{k}$). Vector $\mathbf{k}$ modifies the capacity constraints. Once found, the standard distributed NUM algorithm can be run, realizing thus (1). Associated with the solution of (2) (and obtained via the distributed algorithm), is a vector of Lagrange multipliers $\boldsymbol{\lambda}(\mathbf{k}) = (\lambda_l(\mathbf{k}), l \in \mathcal{L})$. Using a price-based control of user demand framework, we assume that users maximise their surplus. Hence if for flow i the price per unit of bandwidth is p_i, the user will choose demand f_i, such that

$$f_i = \arg\max_f \{U_i(f) - p_i f\}. \quad (3)$$

In this setting, and for concave utility functions, surplus is maximized at a point where marginal utility equals price. When the price seen by flow i is given by the sum of Lagrange multipliers $\boldsymbol{\lambda}(\mathbf{k})$ over all links on path R_i,

$$p_i(\mathbf{k}) = \sum_{l \in R_i} \lambda_l(\mathbf{k}), \quad (4)$$

then the solution f_i found by each user corresponds to the solution $f_i^*(\mathbf{k})$ of NUM($\mathbf{k}$) [3]. The value of the revenue indicator π (referred to as revenue from now on) is calculated as if the operator charged prices $p_i(\mathbf{k})$ for successfully carried traffic,

$$\pi(\mathbf{k}) = \sum_{i \in \mathcal{F}} p_i(\mathbf{k}) f_i^*(\mathbf{k})(1 - L^{(i)}), \quad (5)$$

where $L^{(i)}$ denotes the loss encountered by flow i along its path.

Manuscript received February 23, 2007. The associate editor coordinating the review of this letter and approving it for publication was Dr. Aleksandra Smiljanic. The work of D. Mayer was supported by the UK Engineering and Physical Sciences Research Council under Grant GR/S52360/01.

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK (e-mail: {david.mayer, j.barria}@imperial.ac.uk).

Digital Object Identifier 10.1109/LCOMM.2007.070280.

We use a single-server queue with a buffer of size K to model link loss in the network. Note that our algorithm is not dependent on the specific loss function. We give numerical results for both small and large buffer, accounting for different loss behavior. We assume that packets arrive to link l according to a Poisson process with mean rate x_l and the packet length is exponentially distributed with mean s. Service rate is then exponentially distributed with mean $\mu_l = C_l/s$. This is an M/M/1/K queue, for which the loss probability is defined as $L_l(x_l) = \rho_l^K \frac{1-\rho_l}{1-\rho_l^{K+1}}$, where $\rho_l = \frac{x_l}{\mu_l} = \frac{x_l}{C_l}s$ is the offered load. Furthermore, assuming link independence, the loss encountered by flow i can be calculated as $L^{(i)} \approx 1-\prod_{l\in R_i}(1-L_l)$, where $L_l, l \in \mathcal{L}$ is obtained from a system of fixed-point equations as in [6].

III. Proposed Solution

Problem (1) has a concave objective function, but nonconvex constraints (revenue (5) is a non-monotonic function of demand) for which a suitable optimization method is that of Augmented Lagrangians [7]. This is essentially a penalty method, which converts constrained problems into a sequence of unconstrained problems $Q_{\gamma_t}(\cdot), t = 0, 1, \ldots$ by adding a penalty function $P_\gamma(\cdot)$ and an increasing penalty factor γ. The solution of $Q_{\gamma_t}(\cdot)$ is used as the starting point for solving $Q_{\gamma_{t+1}}(\cdot)$. This method does not require a feasible initial point and is well-suited for nonlinear constraints.[1] The augmented Lagrangian function is

$$Q_\gamma(\mathbf{k}, \mu) = U(\mathbf{k}) - P_\gamma(\mathbf{k}, \mu). \tag{6}$$

The inclusion of Lagrange multiplier μ within the penalty function enables convergence without the penalty factor $\gamma \to \infty$.[2] We use the following penalty (adopted from [7])

$$P_\gamma(\mathbf{k}, \mu) = \frac{1}{2\gamma}\left([\max\{0, \mu + \gamma g(\mathbf{k})\}]^2 - \mu^2\right), \tag{7}$$

where $g(\mathbf{k}) = \pi_{min} - \pi(\mathbf{k})$. The maximization of the unconstrained Lagrangian (6) requires the sensitivity of total utility and revenue generated by the solution of (2) with respect to capacity multipliers $\mathbf{k}$. From the theory of Lagrange multipliers it is known that sensitivity of the objective function with respect to constraints equals the corresponding Lagrange multipliers, here $\frac{\partial U}{\partial C_j} = \lambda_j$, hence $\frac{\partial U}{\partial k_j} = C_j\lambda_j$. Note that the sensitivity of revenue $\frac{\partial \pi}{\partial k_j}$ has to be replaced by a finite difference $\frac{\Delta\pi}{\Delta k_j}$ due to discontinuities of Lagrangians caused by the change of the set of active constraints in (2). Sensitivity of revenue is derived below by extending the results in [5] for interdependent demand and price.

The proposed optimization consists of repeated increases of the augmented Lagrangian (6) by shifting vector $\mathbf{k}$ in the steepest ascent direction,

$$\mathbf{k}_{t+1} = \mathbf{k}_t + \alpha \cdot \nabla Q_\gamma(\mathbf{k}_t, \mu_t), \tag{8}$$

[1]Convergence to global maximum requires that in each iteration a global maximum of $Q(\cdot)$ is found. We rely on the initial point lying in proximity of the global maximum. This is in practice, given by common packet loss functions, $\mathbf{k} = 1$, i.e. the point corresponding to the original link capacities.

[2]The idea is to increase the penalty factor to approach feasible range and then to adjust only the Lagrangian multiplier in the near-optimum range, preventing $\gamma \to \infty$.

where α is a decreasing step size. The gradient of (6) at iteration t is

$$\nabla Q_{\gamma_t}(\mathbf{k}_t, \mu_t) = \mathbf{C}\cdot\lambda(\mathbf{k}_t) + \max\{0, \mu_t + \gamma_t g(\mathbf{k}_t)\}\cdot\frac{\Delta\pi(\mathbf{k}_t)}{\Delta k}. \tag{9}$$

Penalty factor γ and multiplier μ are updated as in [7, p.123]:

$$\mu_{t+1} = max\{0, \mu_t + \gamma_t(\pi_{min} - \pi(\mathbf{k}_t))\} \tag{10}$$

$$\gamma_{t+1} = \begin{cases} \beta\gamma_t & \text{if } \frac{|\pi_{min}-\pi(\mathbf{k}_t)|}{|\pi_{min}-\pi(\mathbf{k}_{t-1})|} \geq \delta \\ \gamma_t & \text{otherwise.} \end{cases} \tag{11}$$

Scalar β is typically chosen $\beta \in [1.1, 2]$ and $\delta = 0.75$. The algorithm terminates when the following condition holds:

$$|\pi_{min} - \pi| \leq \epsilon_\pi \;\wedge\; \|\nabla Q_{\gamma_t}(\mathbf{k}_t, \mu_t)\| \leq \epsilon_P, \tag{12}$$

where typically $\epsilon_\pi = 0.01$ and $\epsilon_P = 1$. The pseudocode of the proposed algorithm for solving (1) is shown in Table I.

Sensitivity of revenue to capacity constraints

In order to update the capacity multipliers vector $\mathbf{k}$ in (8) we need to obtain the sensitivity of revenue with respect to $\mathbf{k}$, $\frac{\Delta\pi}{\Delta k} = (\frac{\Delta\pi}{\Delta k_j}, j = 1, 2, \ldots, L)$. Let us write revenue (5) as a function of demand and price, $\pi = y(\mathbf{f}, \mathbf{p})$. First, we derive the total derivative with respect to demand f_i,

$$\frac{\mathrm{d}\pi}{\mathrm{d}f_i} = \frac{\partial\pi}{\partial f_i} + \frac{\partial\pi}{\partial p_i}\frac{\partial p_i}{\partial f_i}. \tag{13}$$

Revenue sensitivity is known [5] and can be calculated as $\frac{\partial\pi}{\partial f_i} = (1 - L^{(i)})(p_i - \sum_{l\in R_i} c_{il})$, where c_{il} is the implied cost of flow i on link l calculated using the expected buffer occupancy of an M/M/1/K queue. From revenue definition (5) we have $\frac{\partial\pi}{\partial p_i} = f_i(1 - L^{(i)})$. From (3) we have $p_i = U_i{}'$, hence $\frac{\partial p_i}{\partial f_i} = U_i{}''(f_i)$. Then (13) can be expressed as

$$\frac{\mathrm{d}\pi}{\mathrm{d}f_i} = (1 - L^{(i)})\left(p_i - \sum_{l\in R_i} c_{il} + f_i U_i^{''}(f_i)\right). \tag{14}$$

The term $f_i U_i^{''}(f_i)$ is negative and accounts for the dependency of price and demand. We use (14) in the calculation of the revenue sensitivity, which is given by the following total finite difference

$$\frac{\Delta\pi}{\Delta k_j} = C_j\frac{\Delta\pi}{\Delta C_j} = C_j\sum_{i\in\mathcal{F}}\frac{\mathrm{d}\pi}{\mathrm{d}f_i}\frac{\Delta f_i}{\Delta C_j}. \tag{15}$$

Sensitivity of demand to capacity constraints

The remaining unknown in (15) is the change of demand Δf_i upon adding a small ΔC to link j in (2), $\frac{\Delta f_i}{\Delta C_j}$, which we approximate as follows. Total utility U is given by a sum of individual utilities, hence it can be approximated by a sum of Taylor series approximations, here of second order:

$$U(\mathbf{f}^* + \mathbf{\Delta f}) \approx U(\mathbf{f}^*) + \sum_{i\in\mathcal{F}} U_i'\Delta f_i + \sum_{i\in\mathcal{F}}\frac{1}{2}U_i''(\Delta f_i)^2, \tag{16}$$

where $\mathbf{f}^*$ is the current utility maximising demand vector and $\mathbf{\Delta f}$ is the vector of demand changes. Vector $\mathbf{\Delta}_j\mathbf{f}^*$ of optimal changes upon adding small ΔC to link j is approximated by the solution of the following optimization:

$$\max_{\mathbf{\Delta}_j\mathbf{f}} \quad \sum_{i\in\mathcal{F}}\left[U_i'(f_i^*)\Delta_j f_i + \frac{1}{2}U_i''(f_i^*)(\Delta_j f_i)^2\right] \tag{17}$$

$$\text{s.t.} \quad x_l + \sum_{i\in\mathcal{F}}\Delta f_i \cdot a_{i,l} \leq \begin{cases} C_l, l \in \mathcal{L}\backslash j \\ C_l + \Delta C, l = j \end{cases}$$

TABLE I

Proposed algorithm	First iteration for network in Fig. 1a
1. Choose minimum revenue π_{min}.	$\pi_{min} = 5.11$
$t \leftarrow 0$; $\mathbf{k}_t \leftarrow 1$;	$\mathbf{k}_0 = (1, 1)$
2. Obtain demands $\mathbf{f}^*(\mathbf{k}_t)$ and Lagrangians	$\mathbf{f}^* = (0.00, 3.00, 2.00)$
$\boldsymbol{\lambda}(\mathbf{k}_t)$ from the solution of NUM($\mathbf{k}_t$) (2).	$\boldsymbol{\lambda}= (0.91, 1.05)$
3. Calculate revenue π from (5).	$\pi = 4.82$
Evaluate $\nabla U = C\boldsymbol{\lambda}(\mathbf{k}_t)$ and $\Delta\pi$ (15).	$\nabla U = (1.83, 3.15)$
If the termination condition (12) holds,	$\Delta\pi = (0.01, -4.75)$
then $\mathbf{k}^* \leftarrow \mathbf{k}_t$ and TERMINATE.	
4. Update multiplier μ_t (10) and then	$\mu_0 = 0.00$ $\mu_1 = 1.38$
penalty factor γ_t (11).[4]	$\gamma_0 = 4.33$ $\gamma_1 = 4.76$
5. $t \leftarrow t + 1$. Update vector $\mathbf{k}_t$ (8).	$\mathbf{k}_1 = (1.00, 0.98)$
GO TO STEP **2**.	

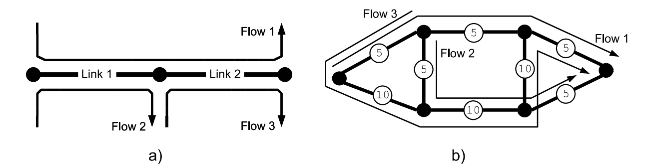

Fig. 1. Networks used in the numerical examples. Circles in network b) specify link capacities.

The rationale for this optimization is that the sum of utility changes $\sum_{i\in\mathcal{F}} \Delta U_i$ approximated by the last two terms in (16) must be maximal and that link flow must not exceed any link capacity except for the augmented link j. This optimization has the standard form of a quadratic objective function $\boldsymbol{\Delta}_j\mathbf{f}\ \mathbf{y} + \frac{1}{2}\boldsymbol{\Delta}_j\mathbf{f}\ \mathbf{H}\ (\boldsymbol{\Delta}_j\mathbf{f})^T$ where, in our case, $\mathbf{y} = (U_1'(f_1^*), \ldots, U_F'(f_F^*))^T$, $\mathbf{H}$ is a diagonal matrix of second derivations with $H(i,i) = U_i''(f_i^*)$ and so (17) can be easily solved using standard algorithms (e.g. [7]). Solution $\boldsymbol{\Delta}_j\mathbf{f}^*$ of (17) is used to approximate $\frac{\Delta f_i}{\Delta C_j}$ in (15).

IV. Numerical Results

First we show the proposed procedure on a 2-link network accommodating 3 users (see Fig. 1a). Utility functions of the users are in the form $U_i(f_i) = a_i(1 - 2^{-b_i f_i})$ and parameters of the problem are: $a_1 = 2, a_2 = 5, a_3 = 10, b_1 = b_2 = b_3 = 0.8, C_1 = 2, C_2 = 3$. Solutions for various revenue constraints are shown in Fig. 2, where axes represent elements of vector $\mathbf{k}$. Contours of revenue and total utility were obtained by exhaustive search. All solutions were reached starting from $\mathbf{k}_0 = (1, 1)$, although in practice previous solutions could serve as a starting point. Table I shows the data for the first iteration of the algorithm for $\pi_{min} = 5.11$. Fig. 3a shows the revenue-utility trade-off for various buffer size to demonstrate the dependence of the trade-off on loss characteristics of the system.[3] Results for a larger 8-link network (see Fig. 1b) are shown in Fig. 3b.

V. Final Remarks

This letter addresses network utility maximization with a constraint on an operator's revenue indicator and proposes an

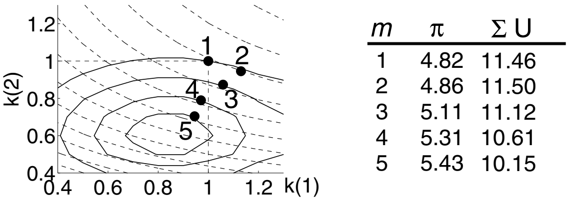

Fig. 2. Contours of revenue (full line) and utility (dashed line) for network in Fig. 1a. Circles show solutions for particular revenue constraints π_{min} indexed by m. Solutions were reached individually from starting point $\mathbf{k}_0 = (1, 1)$ using the proposed algorithm and for buffer size $K = 24$.

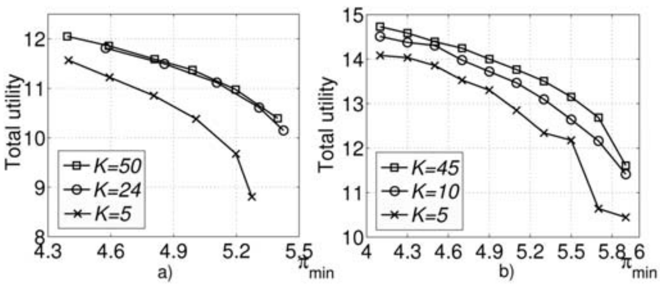

Fig. 3. Total utility as a function of minimum revenue constraint for different buffer sizes K as obtained by the proposed algorithm for network in Fig. 1a (a) and Fig. 1b (b).

algorithm for controlling this trade-off. Although some parameters of the proposed algorithm are computed centrally (the revenue sensitivity vector), once they are communicated to network nodes, the NUM algorithm runs in a fully distributed way. At present we are developing the distributed version of this framework which will be reported in a forthcoming paper.

Regarding the assumption of concave utility functions, recent work on maximization of non-concave utility functions offers new methods based on the standard NUM algorithm. We are therefore also extending our work in this direction.

References

[1] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker, "On selfish routing in Internet-like environments," in *Proc. SIGCOMM03*, Karlsruhe, Germany, 2003.

[2] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, 1998.

[3] S. Low and D. Lapsely, "Optimization flow control. I. Basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, no. 6, pp. 861–874, 1999.

[4] M. Zukerman, L. Tan, H. Wang, and I. Ouveysi, "Efficiency-fairness tradeoff in telecommunications networks," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 643–645, 2005.

[5] D. Mitra, J. Morrison, and K. Ramakrishnan, "ATM network design and optimization: a multirate loss network framework," *IEEE/ACM Trans. Networking*, vol. 4, no. 4, pp. 531–543, 1996.

[6] F. Kelly, "Loss networks," *The Annals of Applied Probability*, vol. 1, no. 3, pp. 319–378, 1991.

[7] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, INC. (LONDON) LTD., 1982.

[3] Large buffer size corresponds to effective bandwidth close to the average rate (small loss), whereas small buffer size models effective bandwidth close to the peak rate (high loss).

[4] Multipliers are initialized as $\mu_0 = 0$ and $\gamma_0 = \left|\frac{\|\nabla U\|}{\|\nabla\pi\|\cdot(\pi_{min}-\pi)}\right|$.