

# Joint QoS Routing and Dynamic Capacity Dimensioning with Elastic Traffic: A Game Theoretical Perspective

Jocelyne Elias  
University of Bergamo  
E-mail: jocelyne.elias@unibg.it

Fabio Martignon  
University of Bergamo  
E-mail: fabio.martignon@unibg.it

**Abstract**—Efficient dynamic resource provisioning algorithms are necessary to the development and automation of Quality of Service (QoS) networks. The main goal of these algorithms is to offer services that satisfy the QoS requirements of individual users while guaranteeing at the same time an efficient utilization of network resources.

This paper proposes a novel game theoretical model that solves the joint problem of non-cooperative QoS routing and dynamic capacity allocation in a parallel links network. Two categories of players are introduced: (1) the *capacity players* that dimension the link capacities to provide QoS guarantees to users, minimizing, at the same time, the links' congestion, and (2) *network users*, which are characterized by elastic traffic demands and split their traffic over multiple links, maximizing their objective function. This game is modeled as a multi-leader-follower game, where capacity players are leaders and network users are followers.

We derive optimal routing and capacity settings using a round robin greedy algorithm, discussing numerical examples that provide insights into the model's solution.

**Index Terms:** - QoS Routing, Capacity Dimensioning, Multi-Leader-Follower Game, Elastic Traffic.

## I. INTRODUCTION

Efficient dynamic resource provisioning mechanisms are necessary to the development and automation of Quality of Service (QoS) networks.

In communication networks, resource allocation is performed mainly in a static way, on time scales on the order of hours to months. However, statically provisioned network resources can become insufficient or considerably under-utilized if traffic statistics change significantly [1]. Telecommunication companies often overbook network resources to increase their income, but overbooking is often performed statically, without considering users' behavior and perceived utility [2].

Therefore, a key challenge for the deployment of Quality of Service networks is the development of solutions that can dynamically track traffic statistics and allocate network resources efficiently, satisfying the QoS requirements of users while maximizing, at the same time, resource utilization and network revenue.

This work deals with the joint problem of QoS routing and dynamic capacity allocation in parallel links networks (illustrated in Figure 1) using Game Theory. Such networks constitute an appropriate model for several networking problems:

(a) modern networking, where resources are pre-allocated to various non-interfering routing paths, (b) broadband networks, where the bandwidth is partitioned among different virtual paths, (c) internetworking, where each link represents a different sub-network.

We consider two categories of players:

- 1) *Capacity Players* (CPs), which own network links, and dimension dynamically the links' capacities to satisfy the QoS requirements of network users.
- 2) *Network users*, which compete in a non-cooperative way for the available resources, maximizing their objective functions (which are related to their individual throughput). In our work, we consider users characterized by elastic traffic demands, like for TCP traffic, which constitutes the majority of the Internet traffic.

As a consequence, Game Theory is the natural framework to study the interactions among such players.

Competitive routing with selfish users has been the focus of several works related to telecommunication networks [3], [4], [5]. Very few works have considered the joint problem of non-cooperative routing and capacity allocation in communications networks, especially in parallel links networks [6], [7]. However, these works present the following limitations:

- 1) in [6], [7], users have a fixed amount of traffic to ship between the source and the destination nodes, while, in our work, users are characterized by elastic traffic demands; the amount of traffic transmitted into the network depends therefore on the user perceived utility in sending his traffic and the disutility due to link congestion.
- 2) No service differentiation is envisaged in [7], and capacity allocation is performed off-line during the network design phase.

This paper overcomes the above limitations by proposing a novel game theoretic model that solves the joint problem of QoS routing and capacity allocation, considering both multiple CPs and a finite set of network users. This game is modeled as a multi-leader multi-follower game [8], [9], where CPs are the leaders who dimension dynamically links capacities in order to provide QoS guarantees to users while minimizing, at the same time, network congestion; the followers are the network

users who are characterized by elastic traffic demands that can be shipped over one or multiple links to maximize their objective function.

Then, we illustrate an iterative round robin greedy algorithm that allows both capacity players and network users to converge to a Nash Equilibrium Point (NEP). Finally, we discuss numerical network examples that provide insights into the model's solution.

The paper is structured as follows: Section II introduces the joint QoS routing and capacity allocation model, including the CPs' and users' objective functions, as well as the adopted congestion cost functions. Section III presents the method used to compute a global solution for our problem. Section IV discusses numerical examples that illustrate how our model captures interesting aspects of the interaction between CPs and network users. Finally, Section V concludes this paper.

## II. NETWORK MODEL

The adopted model is a *network of parallel links*, illustrated in Figure 1, with a set  $\mathcal{I} = \{1, \dots, I\}$  of users that share a set  $\mathcal{L} = \{1, \dots, L\}$  of communication links connecting source node  $S$  to destination node  $D$ .

We assume that each link is owned by an operator (which we refer to as "capacity player") and is characterized by a fixed overall capacity to be shared among the set of supported Classes of Services (CoSs), denoted by  $\mathcal{P} = \{1, \dots, P\}$ . The overall capacity  $c_l$  of link  $l$  is partitioned into a portion  $c_l^a$  statically assigned by the  $l$ -th capacity player to the  $P$  CoSs, and in a portion  $c_l^s$  (the *spare* link capacity) partitioned dynamically between the  $P$  CoSs taking into account the traffic conditions and link congestion; clearly,  $c_l = c_l^a + c_l^s$ . In our approach, such a dynamic capacity assignment is performed in a completely non-cooperative setting.

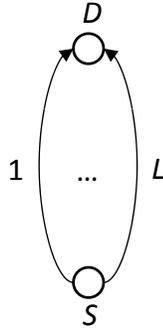


Fig. 1. Parallel Links Network: a source node  $S$  and a destination node  $D$  are interconnected by a set of  $L$  parallel communications links, each owned and managed by a distinct capacity player. A set of  $I$  users need to ship their flow from  $S$  to  $D$ .

User  $i$  ships his flow by splitting it over the set of parallel links, according to some individual performance objective. Users demands are *elastic*, in the sense that they are function of the user perceived utility as well as of the costs due to link congestion. Without loss of generality, we assume that each user  $i$  is characterized by an upper bound ( $M_p^i$ ) and

a lower bound ( $m_p^i$ ) on his demand per class of service  $p$ , which represent his per CoS maximum and minimum traffic requirements, respectively.

Let  $f_{l,p}^i$  denote the flow, belonging to class  $p$ , that user  $i$  sends on link  $l$ . The user flow configuration  $f^i = \{f_1^i, \dots, f_L^i\}$  is called a routing *strategy* of user  $i$ , where each component is a  $P$ -dimensional vector  $f_l^i = (f_{l,1}^i, \dots, f_{l,P}^i) \in R^P$ , and the set of strategies  $F^i = \{f^i \in R^{PL} : f_{l,p}^i \geq 0, l \in \mathcal{L}, p \in \mathcal{P}\}$  is called the routing strategy space of user  $i$ . The system flow configuration  $f = \{f^1, \dots, f^I\}$  is called a routing *strategy profile* and takes values in the product strategy space  $F$ . Furthermore, let  $f^{-i}$  represent the flow configuration of all users except user  $i$ .

Each individual user  $i \in \mathcal{I}$  aims at maximizing his degree of satisfaction (his objective function  $OF_i^u$ , defined in the following), which we assume has a component dependent on the throughput (*utility*) and another one related to costs (*disutility*).

On the other hand, we assume that a group  $I^{CP} = \{1, \dots, L\}$  of *capacity players* participates in the game. We associate a capacity player with each one of the  $L$  links. Capacity players behave selfishly, and each one aims at minimizing the overall congestion over the associated link by allocating appropriately the spare capacity among the  $P$  CoSs. Each capacity player defines his own *capacity strategy*  $c_l^s = (c_{l,p}^s) \in R^P$ , where the quantity  $c_{l,p}^s$ , with  $p \in \mathcal{P}$ , represents the additional capacity assigned to the class  $p$  over link  $l$ . The collection of the capacity players strategies builds up the capacity strategy profile  $c^s = (c_1^s, c_2^s, \dots, c_L^s) \in R^{PL}$ .

The vector  $(f, c^s) \in R^{PL+PL}$  grouping both the routing strategy profile and the capacity strategy profile is referred to as network strategy profile, and represents the solution of the game where both the individual users and the capacity players operate simultaneously.

### A. Multi-Leader-Follower Game

The problem of joint QoS routing and dynamic capacity dimensioning addressed in this work can be solved considering a multi-leader-follower game, which is a general case of a Stackelberg game [10] in which there are multiple leaders and multiple followers who react to the leaders' strategies.

Capacity players are the leaders of the game, and they set the spare capacities independently, since we assume that a capacity player's strategy set is independent of its rivals' strategies. Each user responds to such capacity setting by splitting his flow over the  $L$  links in order to optimize his own objective function. This is naturally a two-stage multi-leader-follower game, in which capacity players act as leaders, with  $I$  users acting as followers that play a non-cooperative game with the spare capacities  $c_l^s$  ( $l = 1, \dots, L$ ) as capacity player strategies.

Similarly to the analysis of a Stackelberg game [10], we first analyze the second stage of this game, which we refer to as *QoS routing sub-game*, assuming that capacity player  $l$  has set its spare capacity  $c_l^s$  in the first stage of the game. All users in the QoS routing sub-game share the same objective

function, and therefore they split their flows on the  $L$  links in order to maximize their profit.

In summary, the QoS routing sub-game takes as input the spare capacities set by capacity players (i.e.,  $c^s = (c_1^s, c_2^s, \dots, c_L^s) \in R^{PL}$ ), and gives as outcome the optimal flow transmitted by each user  $i$  on each link  $l$  for every class of service  $p$ ,  $f_{l,p}^{i*}$ . The flow  $f_{l,p}^{i*}$  is the best response of user  $i$  to the capacity settings  $c^s$ .

In the first stage, capacity players do not cooperate among themselves since they are competing with each other for traffic demand accommodation. They need to set their capacities (in a one-shot game) simultaneously and independently, which implies that the first stage game is a standard Nash game. The action set of capacity player  $l$  in this non-cooperative game is  $c_{l,p}^s > f_{l,p} - c_{l,p}^a$ , where  $f_{l,p}$  is the total amount of flow belonging to class  $p$  on link  $l$ , and  $c_{l,p}^a$  is the static capacity of link  $l$  reserved to class  $p$ .

### B. User Objective Function

We associate to network user  $i \in \mathcal{I}$  the objective function  $OF_i^u$ , which is a function of the flow routed in the network as well as of the congestion costs experienced on the links:

$$OF_i^u(f^i, f^{-i}, c^s) = \sum_{p \in \mathcal{P}} Q_p^i(f_p^i) - \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}} f_{l,p}^i D_{l,p}(f_{l,p}, c_{l,p}^s) \quad (1)$$

$$D_{l,p}(f_{l,p}, c_{l,p}^s) = \begin{cases} \frac{1}{c_{l,p}^a + c_{l,p}^s - f_{l,p}} & , \text{ if } f_{l,p} < c_{l,p}^a + c_{l,p}^s \\ \infty & , \text{ otherwise} \end{cases} \quad (2)$$

where  $D_{l,p}(f_{l,p}, c_{l,p}^s)$  represents the delay experienced by a unit of flow belonging to class  $p$  over link  $l$ .

The first term,  $\sum_{p \in \mathcal{P}} Q_p^i(f_p^i)$ , represents the total utility perceived by the user for routing a total amount of flow  $\sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}} f_{l,p}^i = f^i$ , and we assume that  $Q_p^i(f_p^i)$  is continuous, concave and continuously differentiable in  $f_{l,p}^i, \forall l \in \mathcal{L}, p \in \mathcal{P}$ . The second term,  $\sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}} f_{l,p}^i D_{l,p}(f_{l,p}, c_{l,p}^s)$ , is the disutility due to link delay.

Hence, each user  $i$  aims to maximize his objective function  $OF_i^u$ :

$$\max_{f^i} OF_i^u(f^i, f^{-i*}, c^{s*}) \quad (3)$$

subject to

Non-negativity constraints:

$$f_{l,p}^i \geq 0, \forall l \in \mathcal{L}, p \in \mathcal{P} \quad (4)$$

Demand upper and lower bound constraints:

$$m_p^i \leq \sum_{l \in \mathcal{L}} f_{l,p}^i \leq M_p^i, \forall p \in \mathcal{P} \quad (5)$$

where  $f^{-i*}$  are the optimal flows of all users  $j \in \mathcal{I}$  with  $j \neq i$ , and  $c^{s*}$  are the optimal spare capacities set by all capacity players  $l \in \mathcal{L}$ .

Note that we can consider either general or special users' utility functions. In this work, we assume that each user  $i \in \mathcal{I}$  is characterized by the following utility function:

$$Q^i(f^i) = \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}} \alpha_{l,p}^i f_{l,p}^i \quad (6)$$

where  $\alpha_{l,p}^i$  represents the utility of user  $i$  per unit of flow of class  $p$  on link  $l$ .

Hence, each user  $i$  aims to maximize the following objective function:

$$\max_{f^i} \left\{ \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}} \alpha_{l,p}^i f_{l,p}^i - f_{l,p}^i D_{l,p}(f_{l,p}, c_{l,p}^s) \right\} \quad (7)$$

subject to constraints (4)-(5).

### C. Capacity Player Objective Function

The objective function of the  $l$ -th capacity player,  $OF_l^{CP}$ , is given by:

$$OF_l^{CP}(f, c^s) = \sum_{p \in \mathcal{P}} D_{l,p}(f_{l,p}, c_{l,p}^s) \quad (8)$$

where  $D_{l,p}(f_{l,p}, c_{l,p}^s)$  is the same as in (2).

Each capacity player  $l$  aims to minimize its cost function  $OF_l^{CP}$ :

$$\min_{c_l^s} OF_l^{CP}(f^*, c_l^s, c_{-l}^{s*}) \quad (9)$$

where  $f^*$  and  $c_{-l}^{s*}$  are, respectively, the optimal flow configurations of all users and the optimal spare capacity values of all capacity players except CP  $l$ .

## III. SOLUTION APPROACH

It can be shown that in our proposed game, the NEP exists and is unique. The demonstration is not reported due to space constraints.

We propose to apply a *Round Robin Greedy Algorithm* to solve our proposed game, in order to reach the NEP. In this approach, at each step  $k$ , capacity players modify their capacity strategies ( $c_l^s$ ), and then users take routing decisions ( $f^i$ ).

Let us consider, for the sake of clarity, a network scenario with 3 users and 2 capacity players. At each step  $k$ , the 5 players modify their strategies in a round robin manner according to the following procedure:

$$c_1^{s,k+1} = \arg \min_{c_1^s} OF_1^{CP}(f^{1,k}, f^{2,k}, f^{3,k}, c_1^s, c_2^{s,k})$$

$$c_2^{s,k+1} = \arg \min_{c_2^s} OF_2^{CP}(f^{1,k}, f^{2,k}, f^{3,k}, c_1^{s,k+1}, c_2^s)$$

$$f^{1,k+1} = \arg \max_{f^1} OF_1^u(f^1, f^{2,k}, f^{3,k}, c_1^{s,k+1}, c_2^{s,k+1})$$

$$f^{2,k+1} = \arg \max_{f^2} OF_2^u(f^{1,k+1}, f^2, f^{3,k}, c_1^{s,k+1}, c_2^{s,k+1})$$

$$f^{3,k+1} = \arg \max_{f^3} OF_3^u(f^{1,k+1}, f^{2,k+1}, f^3, c_1^{s,k+1}, c_2^{s,k+1}) \quad (10)$$

where  $arg$  denotes the value of the vector ( $c_l^s$  or  $f^i$  for capacity players and network users, respectively) which optimizes the player objective function.

#### IV. NUMERICAL RESULTS

This section analyzes the dynamic and equilibrium behavior of the proposed joint QoS Routing and Dynamic Capacity Dimensioning (QR-DCD) game, comparing its performance with that of a static scheme, which consists in assigning statically the whole capacity of each link ( $c_l$ ) to the  $P$  classes of services.

Each optimization (minimization or maximization) of expressions (10) is performed by solving a constrained optimization problem, using the Matlab Optimization Tool.

We first consider a network scenario with  $L = 2$  links connecting the source node  $S$  to the destination node  $D$ ,  $I = 3$  users and  $P = 2$  CoSs. We assume that the total capacities of links  $l = 1$  and  $l = 2$  are equal to  $c_1 = 10$  Mb/s and  $c_2 = 30$  Mb/s, respectively; moreover,  $m_p^i = 0$  and  $M_p^i$  is sufficiently large not to limit the transmitted flow of any user. The static scheme assigns 70% of the link capacity to CoS 1, while the rest (30%) is assigned to CoS 2. These are the same percentages adopted to share the *static* portion  $c_l^a$  between the two CoSs in the QR-DCD scheme. The links capacities (including static and spare portions), are reported in Table I. The capacity configuration with the static scheme is illustrated in Table II.

TABLE I

CAPACITY ALLOCATION (MB/S) FOR THE DYNAMIC SCHEME (QR-DCD).

Link	Static Portion ( $c_{l,p}^a$ )			Dynamic Portion ( $c_l^s$ )
	$p = 1$	$p = 2$	Tot. per link	
$l1$	4.2	1.8	6	4
$l2$	12.6	5.4	18	12
Tot. per CoS	16.8	7.2	24	16

TABLE II

CAPACITY ALLOCATION (MB/S) FOR THE STATIC SCHEME.

Link	$p = 1$	$p = 2$	Tot. per link
$l1$	7	3	10
$l2$	21	9	30
Tot. per CoS	28	12	40

The three users have different per flow unit utility values per CoS  $p$ : each user  $i$  has  $\alpha_{i,1}^i = 1$  and  $\alpha_{i,2}^i = 4 \forall i \in \mathcal{I}, l \in \mathcal{L}$ .

At the Nash equilibrium, all users transmit the same amount of flow per link and per class of service. For this reason, we have reported the equilibrium flow only for one user in Table III, for both the QR-DCD and static allocation schemes. Moreover, the Table shows also the spare capacity allocated to each CoS by capacity players 1 and 2.

The utility perceived by each user under the proposed QR-DCD scheme is approximately equal to 23 in this case, and it is more than 57% higher than that obtained by the static capacity allocation scheme ( $\approx 14.57$ ).

*Stability Considerations:* Figures 2 and 3 show how users and capacity players update their strategy under the round robin greedy scheme to converge towards the equilibrium flows and spare capacity allocations (already reported in Table III) in the same scenario considered above. Note that a similar trend has been observed for network users with  $p = 2$ , and for capacity player 2 with both classes of services.

It can be observed that the adoption of our round robin algorithm ensures the convergence of the game towards the Nash equilibrium point for different classes of services and considering elastic users.

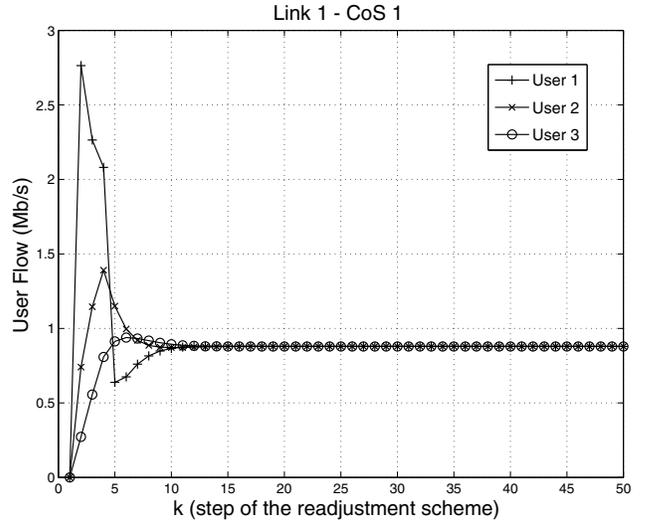


Fig. 2. Convergence to the equilibrium flows for the 3 users on link 1, for  $p = 1$ .

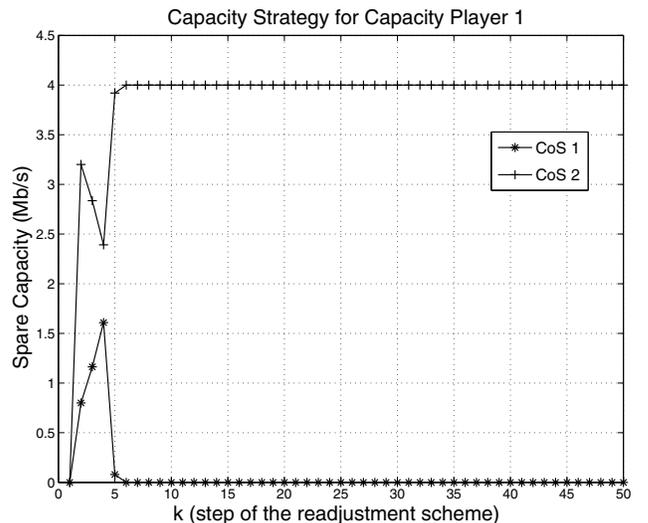
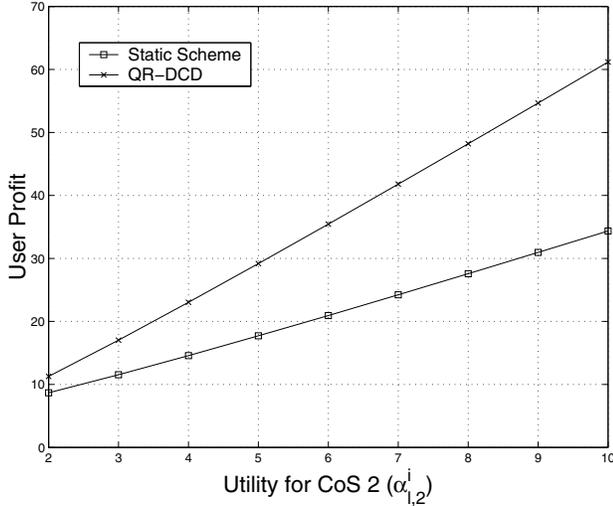


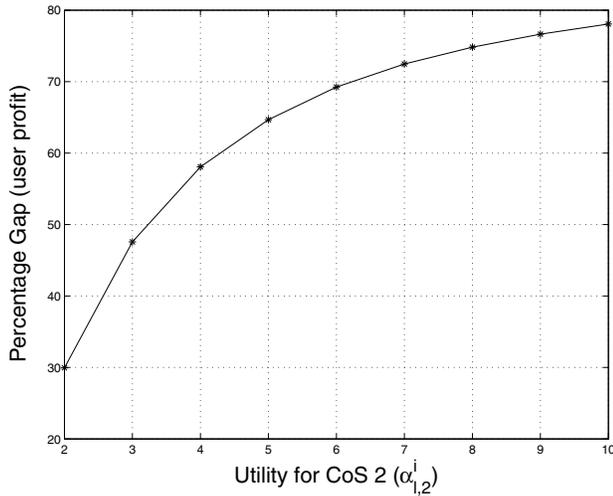
Fig. 3. Convergence to the equilibrium spare capacity allocation for capacity player 1.

We then investigate the effect of different utility values on the equilibrium solution, considering a variation of the previous scenario where each user  $i$  has  $\alpha_{i,1}^i = 1$  and  $\alpha_{i,2}^i$  is varied in the 2 to 10 range  $\forall i \in \mathcal{I}, l \in \mathcal{L}$ .

Figures 4(a) and (b) show, respectively, the profit obtained by each user with the static and QR-DCD scheme, and the percentage gain achieved by QR-DCD over the static scheme in terms of user's profit, when varying the per flow unit utility value ( $\alpha_{1,2}^i$ ).



(a)



(b)

Fig. 4. (a) User profit under the static and QR-DCD scheme, and (b) percentage gain achieved by QR-DCD over the static scheme in terms of user's profit, when varying the per flow unit utility value ( $\alpha_{1,2}^i$ ).

It can be observed that the introduction of capacity players that dimension dynamically the links capacities, taking into account the users' flows, improves consistently the network performance and the users' degree of satisfaction (up to 78% in this scenario). The usefulness of the CPs is much more evident when users have different per flow utility values, as it can be observed in Figure 4.

TABLE III  
DIFFERENT PER FLOW UNIT UTILITY VALUES: EQUILIBRIUM FLOW AND CAPACITY ASSIGNMENTS (MB/S).

Individual User				QR-DCD			
$f_{1,1}^i$	$f_{1,2}^i$	$f_{2,1}^i$	$f_{2,2}^i$	$c_{1,1}^s$	$c_{1,2}^s$	$c_{2,1}^s$	$c_{2,2}^s$
0.88	1.67	3.78	4.97	0.0	4.0	1.25	10.75

Individual User				Static Scheme			
$f_{1,1}^i$	$f_{1,2}^i$	$f_{2,1}^i$	$f_{2,2}^i$	$c_{1,1}^s$	$c_{1,2}^s$	$c_{2,1}^s$	$c_{2,2}^s$
1.70	0.80	6.00	2.68	2.8	1.2	8.4	3.6

## V. CONCLUSION

In this paper we proposed a novel game theoretical model that solves the joint problem of non-cooperative QoS routing and dynamic capacity allocation in a parallel links network.

This problem is solved using a two-stage multi-leader-follower game, where the leaders are the capacity players that dimension the links capacities to provide QoS guarantees to users, minimizing, at the same time, the link's congestion, and the followers are network users, which are characterized by elastic traffic demands and split their traffic over multiple links maximizing their objective function.

We illustrated an iterative round robin greedy algorithm that allows both capacity players and network users to converge in a few steps to a NEP. Finally, we discussed numerical results that illustrate the effectiveness of the proposed solution.

## ACKNOWLEDGMENT

This work was partially supported by MIUR and FP7 in the framework of the PRIN SESAME and PrimeLife projects.

## REFERENCES

- [1] A.T. Campbell and R.R.-F. Liao. Dynamic Core Provisioning for Quantitative Differentiated Services. *IEEE/ACM Transactions on Networking*, pages 429–442, vol. 12, no. 3, June 2004.
- [2] N.G. Duffield, P. Goyal, A. Greenberg, P. Mishra, K.K. Ramakrishnan, and J.E. van der Merwe. Resource Management With Hoses: Point-to-Cloud Services for Virtual Private Networks. *IEEE/ACM Transactions on Networking*, pages 679–692, vol. 10, no. 5, October 2002.
- [3] A. Orda, R. Rom, and N. Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking*, pages 510–521, vol. 1(5), 1993.
- [4] Y.A. Korilis, A.A. Lazar, and A. Orda. Architecting noncooperative networks. *IEEE Journal on Selected Areas in Communications*, pages 1241–1251, vol. 13, no. 7, 1995.
- [5] E. Altman, T. Basar, T. Jimenez, and N. Shimkin. Competitive routing in networks with polynomial costs. *IEEE Transactions on Automatic Control*, pages 92–96, vol. 47(1), 2002.
- [6] P. Conforto and F. Delli Priscoli. Routing and dynamic resource assignment joint game: a non-cooperative model for QoS routing. *International Journal of Control*, pages 1408–1425, vol. 77(16), November 2004.
- [7] Y.A. Korilis, A.A. Lazar, and A. Orda. Achieving network optima using Stackelberg routing strategies. *IEEE/ACM transactions on networking*, pages 161–173, vol. 5, no. 1, 1997.
- [8] H.D. Sherali. A multiple leader Stackelberg model and analysis. *Operations Research*, pages 390–404, 1984.
- [9] J.S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Computational Management Science*, pages 21–56, vol. 2(1), January 2005.
- [10] T. Basar and G. J. Olsder. *Dynamic noncooperative game theory*. SIAM Series in Classics in Applied Mathematics, Philadelphia, PA, 1999.