

An Optimal Congestion and Cost-Sharing Pricing Scheme for Multiclass Services

Yezekael Hayel, Bruno Tuffin

N°4757

Mars 2003

_____ THÈME 1 _____



*apport
de recherche*

An Optimal Congestion and Cost-Sharing Pricing Scheme for Multiclass Services

Yezekael Hayel, * Bruno Tuffin †

Thème 1 — Réseaux et systèmes
Projet Armor

Rapport de recherche n° 4757 — Mars 2003 — 24 pages

Abstract: We study in this paper a social welfare optimal congestion-pricing scheme for multiclass queuing services which can be applied to telecommunication networks. Most of the literature has focused on the marginal price. Unfortunately, it does not share the total cost among the different classes. We investigate here an optimal Aumann-Shapley congestion-price which verifies this property. We extend the work on the Aumann-Shapley price for priority services, based on the results on the marginal price: instead of just determining the cost repartition among classes for *given* rates, we obtain the rates and charges that optimize the social welfare.

Key-words: Pricing, queuing networks, optimization.

(Résumé : *tsvp*)

* yhayel@irisa.fr

† btuffin@irisa.fr

Une méthode de tarification optimale basée sur la congestion et le partage des coûts pour les services multiclassés

Résumé : Nous étudions dans ce rapport une méthode de tarification optimale basée sur la congestion pour des services multiclassés tels que les futurs réseaux de télécommunications. La plupart des modèles pour ce type d'architecture utilise une tarification au coût marginal. Seulement, celui-ci ne réalise pas un partage du coût total entre les classes de service. Nous utilisons le mécanisme de Aumann-Shapley qui possède cette propriété afin de déterminer la part du coût global attribuée à chacune des classes. Enfin, nous étendons le travail en nous basant sur les résultats de la tarification au coût marginal: au lieu de déterminer les prix pour des taux d'arrivées fixés, nous obtenons les taux et les prix qui optimisent le *social welfare*.

Mots-clé : Tarification, réseaux de files d'attente, optimisation.

1 Introduction

During the last decade, the Internet has experienced a tremendous growth of its traffic. The current flat-rate pricing scheme, adopted by most Internet Service Providers (ISPs), is an incentive to overuse the network which drives to congestion and reduces the Quality of Service (QoS). The future network architecture will have to respond to different QoS requirements of different types of applications. Architectures such as DiffServ proposal [?] deal with this problem by assigning different priorities at each node of the network and an adapted pricing scheme has to be associated with, otherwise a user would always choose the service class providing the best QoS. So, using a charging scheme can give rise to an efficient allocation of resources among diverse and self-interested users. Indeed, users of a common facility impose a cost on each other, that is known as the *negative externality*, or *congestion cost* [?]. In general, this cost is based on the delay that is imposed on users prior to the completion of their tasks. Most pricing models use the marginal cost mechanism, which is a fundamental economic principle [?]. But it presents the drawback of not sharing the costs among users [?]. In economic theory, there exists a cost-sharing model called the Aumann-Shapley price mechanism which shares the total cost among participants. This axiomatic approach to cost-sharing can be regarded as an extension of the average cost pricing to the multi-product case [?]. The theory of non-atomic games as developed by Aumann and Shapley was first applied by Billera et al [?] to set fair telephone billing rates that share the cost of service among users. We are going to use this cost-sharing mechanism to allocate congestion costs in multi-class priority services like done by Haviv in [?]. We are going to extend the model of [?] based on the tremendous literature about the marginal cost mechanism. Indeed, in [?], the arrival rates were fixed and the cost repartition among classes was then computed. First, like in [?], we determine the prices and rates maximizing the network's social welfare. In order to do that, we will introduce a function representing how much a given class values its service rate, and a demand relationship relating this rate to the actual charge. Then, we extend our result to a network and to a dynamic pricing scheme, like proposed by Masuda et al. in [?] for the marginal cost.

The paper is organized as follows. In Section 2, we describe the Aumann-

Shapley cost-sharing mechanism and its application to network pricing as proposed by Haviv in [?] for given arrivals rates. Then we extend the results of [?] by applying the ideas used in the marginal cost literature. We first consider in Section 3 the optimal Aumann-Shapley price, as in the Mendelson's approach [?] thanks to the introduction of a valuation function and a demand relationship and we also obtain some incentive compatibility results. We look in Section 4 at the *Incentive Compatibility* property in the heterogeneous service case and present a solution to get it. Then, we study two extensions in Sections 5 and 6: the general network case and the time-dynamic pricing where the user's expectations are supposed to be Markovian (following [?]). The conclusions and directions for future research are given in Section 7.

2 The Aumann-Shapley cost-sharing mechanism applied to network pricing

We first introduce the basic model in this section. Then, in the second step, we present the Aumann-Shapley mechanism applied to this model to share the total expected delay cost.

2.1 Queuing model

The model is a multi-class queue where we assume that arrivals are decomposed into n classes. The arrival process of class- i customers follows a Poisson distribution with rate λ_i , $1 \leq i \leq n$. The processing requirement of a class- i job, τ_i , follows an exponential distribution with finite mean c_i . Processing requirements are independent across jobs and identically distributed within job classes. Each class- i job is characterized by a delay cost of v_i per unit of time representing how he *feels* a unit of delay. The model is based on a non-preemptive priority rule where classes are sorted so that

$$\frac{v_1}{c_1} \geq \frac{v_2}{c_2} \geq \dots \geq \frac{v_n}{c_n}, \quad (1)$$

assigning the highest priority to class 1, the second highest priority to class 2 and so on. Let $W_i(\underline{\lambda})$ be the expected queuing delay of a class- i job when

$\underline{\lambda} = (\lambda_1, \dots, \lambda_n)$ is the vector of arrival rates. We have [?],

$$W_i(\underline{\lambda}) = \frac{\sum_{j=1}^n c_j^2 \lambda_j}{(1 - \sum_{j=1}^{i-1} c_j \lambda_j)(1 - \sum_{j=1}^i c_j \lambda_j)} + c_i. \quad (2)$$

Define by $L_i(\underline{\lambda})$ the expected mean number of class- i jobs in the system. By Little's law, we have

$$L_i(\underline{\lambda}) = \lambda_i W_i(\underline{\lambda}).$$

The idea is to apply the Aumann-Shapley price mechanism to the global cost function $L(\underline{\lambda})$ defined by

$$L(\underline{\lambda}) = \sum_{i=1}^n v_i \lambda_i W_i(\underline{\lambda}) \quad (3)$$

which represents the total delay cost incurred by the jobs in the system. Before doing that, we look in the next subsection at the definition and properties of the Aumann-Shapley cost-sharing mechanism in a general economic context.

2.2 The Aumann-Shapley costs

Consider a general cost function $F(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^+$ with $F(0) = 0$ and where $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ represents some levels of inputs (in full generality). F is supposed to be differentiable and nondecreasing in all its variables. The Aumann-Shapley mechanism is defined by several properties. These axioms characterize this cost-sharing mechanism. The goal is to determine $P_i(F, \underline{x})$, the per-unit cost of level- i inputs associated with cost function F . We have:

Axiom 1 (Cost-Sharing) For every $\underline{x} \in \mathbb{R}^n$,

$$\sum_{i=1}^n x_i P_i(F, \underline{x}) = F(\underline{x}).$$

Note that $x_i P_i(F, \underline{x})$ is class- i 's share of $F(\underline{x})$. The total cost is shared between all input's levels.

Axiom 2 (Additivity) *If we have three cost functions F, G and H , such that $F + G = H$, then,*

$$\forall i, \quad P_i(F, \underline{x}) + P_i(G, \underline{x}) = P_i(H, \underline{x}).$$

Axiom 3 (Positivity) *If F is a cost function, then $\forall \underline{x} \in \mathbb{R}^n$,*

$$\forall i, \quad P_i(F, \underline{x}) \geq 0.$$

Axiom 4 (Consistency) *Let F be a cost function and G be another cost function defined over \mathbb{R} , such that,*

$$F(x_1, \dots, x_n) = G\left(\sum_{i=1}^n x_i\right).$$

Then, for each i , $1 \leq i \leq n$, we have

$$P_i(F, \underline{x}) = P\left(G, \sum_{i=1}^n x_i\right),$$

where P is the per-unit cost associated with cost function G . This means that splitting commodities into irrelevant classifications has no effect on the shares of the total cost.

Axiom 5 (Rescaling) *Let F be a cost function and $(\lambda_1, \dots, \lambda_n)$ be n positive real numbers. Let G be another cost function defined by:*

$$G(x_1, \dots, x_n) = F(\lambda_1 x_1, \dots, \lambda_n x_n).$$

Then, for each i , $1 \leq i \leq n$, we have:

$$P_i(G, \underline{x}) = \lambda_i P_i(F, (\lambda_1 x_1, \dots, \lambda_n x_n)).$$

Thus, changing the scale of a commodity yields an equivalent change in the shares.

One can refer to [?] for different works on these axioms. In the following, we obtain a characterization of the Aumann-Shapley cost.

Theorem 1 [?] *There exists one and only one class- i unit cost mechanism $P_i(.,.)$ which obeys the above five axioms. It is called the Aumann-Shapley class- i unit cost mechanism, and it verifies*

$$P_i(F, \underline{x}) = \int_0^1 \frac{\partial F}{\partial x_i}(t\underline{x})dt, \quad i = 1, \dots, n, \quad (4)$$

for each cost function F which is differentiable, with $F(0) = 0$, and $\underline{x} \in \mathbb{R}^n$.

We also have the following result.

Theorem 2 [?] *$P(.,.)$ is a price mechanism obeying Axioms 2-5 if and only if there is a nonnegative measure μ on $([0, 1], \mathcal{B})$ (\mathcal{B} is the family of all Borel subsets of $[0, 1]$) such that for each cost function F and for each $\underline{x} \in \mathbb{R}^n$,*

$$P_i(F, \underline{x}) = \int_0^1 \frac{\partial F}{\partial x_i}(t\underline{x})d\mu(t), \quad i = 1, \dots, n. \quad (5)$$

This theorem defines a one-to-one mapping from the set of price mechanisms obeying Axioms 2-5 onto the set of nonnegative measures on $([0, 1], \mathcal{B})$.

Remark 1 *Formula (5) asserts that the class- i unit costs are a weighted average of the marginal class- i cost of F along the segment $[0, \underline{x}]$. The weights are given by the measure μ . If the measure happens to be the Lebesgue measure on $[0, 1]$, then we obtain the Aumann-Shapley price mechanism (4) and the Axiom 1 is verified. This cost-sharing mechanism is the uniform average of all marginal costs along the segment $[0, \underline{x}]$.*

Remark 2 *If μ happens to be the atomic probability measure whose whole mass is concentrated at the point $t = 1$, i.e., $\mu(\{1\}) = 1$, the associated price mechanism is the marginal cost price [?]. Thus this price mechanism does not verify the cost-sharing Axiom whereas the Aumann-Shapley does. This result is important because most of congestion pricing schemes are based on the marginal cost externality. The Aumann-Shapley cost-sharing pricing scheme can be a fairer response than the marginal cost one.*

2.2.1 Application to the delay cost

Now, we can apply Equation (4) to the delay cost function L to obtain class- i 's share of total cost $L(\underline{\lambda})$ per unit of arrival rate like in [?]:

$$\begin{aligned}
 P_i(L, \underline{\lambda}) &= \int_0^1 \frac{\partial L}{\partial \lambda_i}(t\underline{\lambda}) dt, \\
 &= \int_0^1 \frac{\partial \sum_{j=1}^n v_j \lambda_j W_j}{\partial \lambda_i}(t\underline{\lambda}) dt, \\
 &= \int_0^1 \left(\sum_{j=1}^n v_j \lambda_j t \frac{\partial W_j}{\partial \lambda_i}(t\underline{\lambda}) + v_i W_i(t\underline{\lambda}) \right) dt, \\
 P_i(L, \underline{\lambda}) &= \sum_{j=1}^n v_j \lambda_j \int_0^1 t \frac{\partial W_j}{\partial \lambda_i}(t\underline{\lambda}) dt + v_i \int_0^1 W_i(t\underline{\lambda}) dt. \tag{6}
 \end{aligned}$$

In the next part, we extend [?] where the repartition was computed for fixed arrival rates and we maximize the network value by determining the optimal arrival rate. Then, thanks to a demand relationship, we obtain the optimal price.

3 On the optimality of Aumann-Shapley cost

As our aim is to maximize a notion of user's satisfaction, we need some informations about the service valuation. Class- i jobs are assumed to have heterogeneous values, represented by a value function $V_i(\lambda_i)$ which specifies the gross value gained by systems users (in the aggregate) per unit of time when the arrival rate of class- i jobs to the system is λ_i . For all i , the value function $V_i(\cdot)$ is assumed to be increasing, differentiable and strictly concave. One way to conceptualize this notion is to relate λ_i and the full price z (as the *charge* plus the *felt* cost) by: $\lambda_i = D_i(z) = (1 - H_i(z))\Lambda_i$ where Λ_i is the maximum potential arrival rate of class- i jobs and $H_i(\cdot)$ is the distribution function of the service valuation [?]. Inverting this function, we have $V'_i(\lambda_i) = D_i^{-1}(\lambda_i)$. Let $\underline{\lambda} = (\lambda_1, \dots, \lambda_n)$. The total expected value of the system is

$$V(\underline{\lambda}) = \sum_{i=1}^n V_i(\lambda_i).$$

The problem is to design a pricing scheme that maximizes the social welfare which is the total expected value $V(\underline{\lambda})$ minus the total expected delay cost $L(\underline{\lambda})$. This decision problem can be formulated as

$$\max_{\underline{\lambda}} (V(\underline{\lambda}) - L(\underline{\lambda})). \quad (7)$$

Let the system administrator set the n -dimensional price vector $\underline{p} = (p_1, \dots, p_n)$, where p_i is the unit-price charged to a class- i job. We assume that the class- i demand relationship is

$$\frac{1}{\lambda_i} \int_0^{\lambda_i} V'_i(u) du = \frac{V_i(\lambda_i)}{\lambda_i} = p_i(\lambda_i) + v_i \int_0^1 W_i(\underline{\lambda}t) dt. \quad (8)$$

This relation means that the unit gross value $\frac{V_i(\lambda_i)}{\lambda_i}$ is equal to the mean expected cost which consists in two parts: a direct payment of p_i per job and a uniform average of delay cost $v_i \int_0^1 W_i(\underline{\lambda}t) dt$ incurred by a class- i job.

The price of a class- i job $p_i = p_i(\lambda_i^*) := p_i^*$ is optimal when $\underline{\lambda}^*$ is solution of the social welfare maximization problem (7). We suppose that the service requirements are *homogeneous*, i.e. $c_1 = \dots = c_n = 1$ for all classes ¹. Following (1) the classes are then sorted such that the highest priority is assigned to the class with the highest delay cost: $v_1 > v_2 > \dots > v_n$.

Theorem 3 *The optimal price per class- i job, $i \in \{1, \dots, n\}$, in the homogeneous case is given by*

$$\begin{aligned} p_i^* &= \sum_{j=1}^n v_j \lambda_j^* \int_0^1 t \frac{\partial W_j}{\partial \lambda_i}(\underline{\lambda}^* t) dt, \\ &= \sum_{j=1}^n v_j \lambda_j^* (-\ln(1 - S_j)) \left(\frac{1}{\lambda_j^* S_j} + \mathbb{1}_{i \leq j} \frac{S_n(\lambda_j^{*2} - S_j^2)}{S_j^2 S_{j-1} \lambda_j^{*2}} + \mathbb{1}_{i \leq j-1} \frac{S_n(\lambda_j^{*2} - S_{j-1}^2)}{S_j S_{j-1}^2 \lambda_j^{*2}} \right) \\ &\quad + \sum_{j=1}^n v_j \lambda_j^* (-\ln(1 - S_{j-1})) \left(-\frac{1}{\lambda_j^* S_{j-1}} + \mathbb{1}_{i \leq j} \frac{S_n}{\lambda_j^{*2} S_{j-1}} + \mathbb{1}_{i \leq j-1} \frac{S_n}{\lambda_j^{*2} S_j} \right) \\ &\quad + \sum_{j=1}^n v_j \lambda_j^* \left(\mathbb{1}_{i \leq j} \frac{S_n}{\lambda_j^* S_j (1 - S_j)} - \mathbb{1}_{i \leq j-1} \frac{S_n}{\lambda_j^* S_{j-1} (1 - S_{j-1})} \right). \end{aligned} \quad (9)$$

with $S_j = \sum_{i=1}^j \lambda_i^*$.

¹The last equality can be assumed without loss of generality

Proof The i th first order condition for the maximization problem (7) is

$$V_i'(\lambda_i) = \sum_{j=1}^n v_j \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda}). \quad (10)$$

Furthermore, using the demand relationship (8), we obtain

$$\begin{aligned} \frac{1}{\lambda_i} \int_0^{\lambda_i} V_i'(u) du &= \int_0^1 V_i'(\lambda_i t) dt = p_i + v_i \int_0^1 W_i(\underline{\lambda} t) dt, \\ &= \int_0^1 \sum_{j=1}^n v_j \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda} t) dt. \end{aligned}$$

Thus, we get the optimal price p_i^* ,

$$p_i^* = \int_0^1 \sum_{j=1}^n v_j \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda}^* t) dt - v_i \int_0^1 W_i(\underline{\lambda}^* t) dt.$$

As $L_i(\underline{\lambda}) = \lambda_i W_i(\underline{\lambda})$, a differentiation with respect to λ_i gives

$$\frac{\partial L_i}{\partial \lambda_i}(\underline{\lambda}) = W_i(\underline{\lambda}) + \lambda_i \frac{\partial W_i}{\partial \lambda_i}(\underline{\lambda}) \quad \text{and} \quad \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda}) = \lambda_j \frac{\partial W_j}{\partial \lambda_i}(\underline{\lambda}) \quad \text{for } j \neq i,$$

hence

$$\begin{aligned} p^* &= \int_0^1 \sum_{j \neq i} v_j \frac{\partial \lambda_j^* W_j}{\partial \lambda_i}(\underline{\lambda}^* t) dt + v_i \int_0^1 \left(W_i(\underline{\lambda}^* t) + \lambda_i^* t \frac{\partial W_i}{\partial \lambda_i}(\underline{\lambda}^* t) \right) dt \\ &\quad - v_i \int_0^1 W_i(\underline{\lambda}^* t) dt, \\ &= \sum_{j=1}^n v_j \lambda_j^* \int_0^1 t \frac{\partial W_j}{\partial \lambda_i}(\underline{\lambda}^* t) dt, \\ &= \sum_{j=1}^n v_j \lambda_j^* \int_0^1 t \left(\frac{1}{(1-tS_j)(1-tS_{j-1})} + \frac{\mathbb{1}_{i \leq j} W_j^q}{1-tS_j} + \frac{\mathbb{1}_{i \leq j-1} W_j^q}{1-tS_{j-1}} \right) dt, \end{aligned}$$

with $W_j^q(\underline{\lambda}^* t) = \frac{t \sum_{i=1}^n \lambda_i^*}{(1-tS_j)(1-tS_{j-1})}$ waiting time without service requirements. By partial fraction decomposition in terms of t , we obtain:

$$\frac{t}{(1-tS_j)(1-tS_{j-1})} = \frac{1}{\lambda_j^*(1-tS_j)} - \frac{1}{\lambda_j^*(1-tS_{j-1})},$$

$$\begin{aligned}\frac{tW_j^q}{1-tS_j} &= \frac{t^2 S_n}{(1-tS_j)^2(1-tS_{j-1})}, \\ &= \frac{S_n}{\lambda_j^* S_j (1-tS_j)^2} + \frac{(\lambda_j^{*2} - S_j^2) S_n}{S_j S_{j-1} \lambda_j^{*2} (1-tS_j)} + \frac{S_n}{\lambda_j^{*2} (1-tS_{j-1})},\end{aligned}$$

and finally,

$$\begin{aligned}\frac{tW_j^q}{1-tS_{j-1}} &= \frac{t^2 S_n}{(1-tS_j)(1-tS_{j-1})^2}, \\ &= -\frac{S_n}{\lambda_j^* S_{j-1} (1-tS_{j-1})^2} + \frac{(\lambda_j^{*2} - S_{j-1}^2) S_n}{S_j S_{j-1} \lambda_j^{*2} (1-tS_{j-1})} + \frac{S_n}{\lambda_j^{*2} (1-tS_j)}.\end{aligned}$$

By a simple integration, we obtain the theorem. ■

The optimal price can be considered as the global uniform average of the *externality* cost.

Remark One can see that in contrast with the cost given by the Aumann-Shapley mechanism (6), we have for a class- i job

$$P_i(L, \underline{\lambda}^*) = p_i^* + v_i \int_0^1 W_i(\underline{\lambda}^* t) dt.$$

The optimal total cost $P_i(L, \underline{\lambda}^*)$ is decomposed into a charge (or price) p_i^* and a global cost $v_i \int_0^1 W_i(\underline{\lambda}^* t) dt$ endured (*felt*) by a class- i job .

Now, we can explore some properties about this cost-sharing model. We assume that the system uses a non-preemptive priority discipline among job classes. We look at the *Incentive Compatibility* property defined by Mendelson and Whang in [?], by the fact that each user chooses his correct priority class. Define the total cost incurred by class- i jobs when it is submitted in priority j by

$$P_i^j(L, \underline{\lambda}^*) = p_j^* + v_i \int_0^1 W_j(\underline{\lambda}^* t) dt. \quad (11)$$

One can see that the cost given by the Aumann-Shapley mechanism is $P_i^i(L, \underline{\lambda}^*)$. A formalization of the property is the following.

Definition 1 A priority-dependent pricing scheme is said to be Incentive Compatible if for all $i \in \{1, \dots, n\}$,

$$i = \operatorname{argmin}_{j \in \{1, \dots, n\}} P_i^j(L, \underline{\lambda}^*),$$

meaning that $\forall j \neq i \quad P_i^i(L, \underline{\lambda}^*) < P_i^j(L, \underline{\lambda}^*)$.

Thus if we have an incentive compatible pricing scheme, all users choose their correct service priority class. That is, the overall optimum is a Nash equilibrium. We can prove that, in this homogeneous context, the optimal pricing scheme (10) is incentive compatible.

Theorem 4 Let \underline{p}^* the optimal price vector given by equation (10). Then, \underline{p}^* is incentive compatible.

Proof Consider a class- i job. We have to verify that for any $k \in \{1, \dots, n\} \setminus \{i\}$

$$P_i(L, \underline{\lambda}^*) < P_i^k(L, \underline{\lambda}^*).$$

We have

$$P_i(L, \underline{\lambda}^*) - P_i^k(L, \underline{\lambda}^*) = p_i^* - p_k^* + v_i \int_0^1 (W_i(\underline{\lambda}^* t) - W_k(\underline{\lambda}^* t)) dt,$$

where p_i^* and p_k^* are given by equation (10), so that

$$\begin{aligned} P_i(L, \underline{\lambda}^*) - P_i^k(L, \underline{\lambda}^*) &= \int_0^1 \left(\sum_{j=1}^n v_j \lambda_j^* t \frac{\partial W_j}{\partial \lambda_i}(\underline{\lambda}^* t) \right) - \left(\sum_{j=1}^n v_j \lambda_j^* t \frac{\partial W_j}{\partial \lambda_k}(\underline{\lambda}^* t) \right) dt \\ &\quad + v_i \int_0^1 (W_i(\underline{\lambda}^* t) - W_k(\underline{\lambda}^* t)) dt, \\ &= \int_0^1 (\hat{p}_i(\underline{\lambda}^* t) - \hat{p}_k(\underline{\lambda}^* t) + v_i (W_i(\underline{\lambda}^* t) - W_k(\underline{\lambda}^* t))) dt, \end{aligned}$$

where $\hat{p}_i = \sum_{j=1}^n v_j \lambda_j^* t \frac{\partial W_j}{\partial \lambda_i}(\underline{\lambda}^* t)$. Mendelson and Whang proved in Theorem 2 of [?] (considering the marginal cost mechanism) that,

$$\forall \lambda, \quad \hat{p}_i + v_i W_i < \hat{p}_k + v_i W_k,$$

for all $k \in \{1, \dots, n\} \setminus \{i\}$. Following their result, by integration of a positive function, we obtain the theorem. \blacksquare

This theorem proves that in the context of homogeneous service requirements, the optimal pricing scheme defined by equation (10) is incentive compatible.

Example We consider the same example than in [?] with two classes and value functions

$$V_1(\lambda_1) = 9\lambda_1 - 10\lambda_1^2 \quad \text{and} \quad V_2(\lambda_2) = 12\lambda_2 - 15\lambda_2^2,$$

defined respectively on $[0, 9/20]$ and $[0, 0.4]$. Let $v_1 = 2$, $v_2 = 1$ and consider the homogeneous case with $c_1 = c_2 = 1$. The expected waiting time for each job (given by Equation (2)) is :

$$\begin{aligned} W_1(\underline{\lambda}) &= \frac{\lambda_1 + \lambda_2}{1 - \lambda_1} + 1, \\ W_2(\underline{\lambda}) &= \frac{\lambda_1 + \lambda_2}{(1 - \lambda_1)(1 - \lambda_1 - \lambda_2)} + 1. \end{aligned}$$

We study the incentive compatibility property on this example. In Table 1, we display the total costs and prices obtained for the marginal cost and the Aumann-Shapley cost-sharing mechanisms. We see that the optimal Aumann-

	Marginal cost	Aumann-Shapley
P_1^1	5.34	3.2604
P_1^2	5.71	3.3514
P_2^2	3.66	1.9597
P_2^1	3.78	1.9976
p_1^*	2.21	0.7345
p_2^*	1.61	0.568

Table 1: Total costs and prices for each classes.

Shapley price is incentive compatible. The global cost is about 1.1415, so the total system cost for the optimal Aumann-Shapley price is also 1.1415, but it is 1.99 for the marginal cost. Indeed, as explained in [?], the marginal cost

can over-estimate the total queuing delay. Thus the externality cost is also over-evaluated.

The proportion of the total cost $P_1^1 + P_2^2$ (as the pricing scheme is incentive compatible) for each class and each model are displayed in Table 2, indicating that the proportions among classes are different for the two mechanisms.

	Marginal cost	Aumann-Shapley
first class	59.3%	62.6%
second class	40.7%	37.4%

Table 2: Proportion of total cost for each class

4 Case of heterogeneous service times

In this Section, we extend the property of incentive compatibility to non-homogeneous service requirements (different values of c_i) as done in [?, Section 3], for the marginal cost mechanism. First, consider the following example which is identical to the previous one but with two different service requirements.

Example Consider a system that has two job classes with $V_1(\lambda_1) = 9\lambda_1 - 10\lambda_1^2$ for $\lambda_1 \in [0, 9/20]$ and $V_2(\lambda_2) = 12\lambda_2 - 15\lambda_2^2$ for $\lambda_2 \in [0, 0.4]$. Let $v_1 = 2$, $v_2 = 1$, $c_1 = 0.1$ and $c_2 = 2$. Then by optimizing the social welfare, we obtain the optimal arrival rates $\lambda_1^* = 0.37$ and $\lambda_2^* = 0.15$. A substitution in (10) yields $p_1^* = 0.52$ and $p_2^* = 2.76$. Thus, we have the following total costs from (11): $P_1^1(L, \underline{\lambda}^*) = 1.53$, $P_1^2(L, \underline{\lambda}^*) = 8.07$, $P_2^2(L, \underline{\lambda}^*) = 5.42$ and $P_2^1(L, \underline{\lambda}^*) = 1.03$. Hence, we can conclude that a class-2 job will prefer priority-class 1. In fact, this result comes from the lower price per job, i.e. $p_1^* < p_2^*$ for the first priority-class with a better rate for these jobs, i.e. $\lambda_1^* > \lambda_2^*$.

This example shows that the previous pricing scheme is not incentive compatible in the non-homogeneous case. To prevent this problem, we define a *Priority- and Time-Dependent* pricing scheme. The model will now condition

the job charge on both its priority class- i and on its actual processing time τ_i . In this configuration, we assume that the rates and the prices are related via the following demand-relationship:

$$\frac{V_i(\lambda_i^*)}{\lambda_i^*} = \mathbb{E}(p_i^{*+}(\tau_i)) + v_i \int_0^1 (W_i^q(\lambda^* t) + c_i) dt, \quad (12)$$

where W_j^q is the waiting time in the system without service for a class- j job. This demand-relationship generalizes Equation (8) in the case where the price depends on the processing time τ_i .

Proposition 1 Let p_i^{+*} be defined by

$$p_i^{+*}(\tau_i) = A'_i \tau_i + 1/2 B' \tau_i^2, \quad (13)$$

with

$$\begin{aligned} A'_i(\lambda) = & \left(\sum_{j=1}^n c_j^2 \lambda_j^* \right) \left[\sum_{k=i}^n v_k \left(\frac{1}{c_k S_k (1 - S_k)} \right. \right. \\ & \left. \left. - \frac{(c_k^2 \lambda_k^{*2} - S_k^2) \ln(1 - S_k)}{S_k^2 S_{k-1} c_k^2 \lambda_k^*} - \frac{\ln(1 - S_{k-1})}{c_k^2 \lambda_k^* S_{k-1}} \right) \right. \\ & \left. + \sum_{k=i+1}^n v_k \left(\frac{1}{c_k S_{k-1} (1 - S_{k-1})} - \frac{(c_k^2 \lambda_k^{*2} - S_{k-1}^2) \ln(1 - S_{k-1})}{S_k S_{k-1} c_k^2 \lambda_k^*} \right. \right. \\ & \left. \left. - \frac{\ln(1 - S_k)}{c_k^2 \lambda_k^* S_k} \right) \right] \end{aligned}$$

and

$$B'(\lambda) = - \sum_{k=1}^n v_k \left(\frac{\ln(1 - S_{k-1})}{S_{k-1}} + \frac{\ln(1 - S_k)}{S_k} \right),$$

with $S_k = \sum_{i=1}^k c_k \lambda_k^*$. This pricing scheme is incentive compatible and optimal.

Proof In [?], it has been proved that the pricing scheme $\hat{p}_i^*(\tau_i) = A_i\tau_i + 1/2B\tau_i^2$ with

$$\begin{aligned} A_i(\lambda) &= \frac{a_i}{\bar{S}_{i-1}\bar{S}_i^2} + \sum_{k=i+1}^I a_k \left(\frac{1}{\bar{S}_{k-1}^2\bar{S}_k} + \frac{1}{\bar{S}_{k-1}\bar{S}_k^2} \right), \\ &= \sum_{k=i}^n v_k \lambda_k^* \left(\sum_{j=1}^n c_j^2 \lambda_j^{*2} \frac{1}{(1-S_{k-1})(1-S_k)^2} \right) \\ &\quad + \sum_{k=i+1}^n v_k \lambda_k^* \left(\sum_{j=1}^n c_j^2 \lambda_j^{*2} \frac{1}{(1-S_{k-1})^2(1-S_k)} \right) \end{aligned}$$

and

$$B = \sum_{k=1}^I \frac{v_k \lambda_k^*}{(1-S_{k-1})(1-S_k)},$$

with $a_i = v_i \lambda_i^* \sum_{k=1}^n c_k^2 \lambda_k^*$, is optimal and incentive-compatible in the non-homogeneous marginal cost case. By a simple integration, it can be checked that

$$A'_i = \int_0^1 A_i(u\lambda) du \quad \text{and} \quad B' = \int_0^1 B(u\lambda) du.$$

We are going to show that

$$p_i^{+*}(\tau_i) = A'_i \tau_i + 1/2B' \tau_i^2$$

is incentive compatible and optimal using the results on the marginal cost. We first prove the optimality of the price mechanism (13), and then we verify the incentive compatibility property.

Optimality We just have to show that:

$$\mathbb{E}(p_i^{+*}(\tau_i)) = p_i^*, \quad \forall i \in \{1, \dots, n\},$$

where p_i^* is the optimal price given by (10). From the definitions of A'_i and B' , we can assure that $p_i^{*+}(t)$ is the uniform average of the priority and time dependent price mechanism \hat{p}_i^{*+} of Mendelson and Whang [?, page 879]

$$p_i^{*+} = \int_0^1 \hat{p}_i^{+*}(u\lambda, t) du.$$

Thus we have

$$\begin{aligned}
 \mathbb{E}(p_i^{+*}(\tau_i)) &= \mathbb{E} \left(\int_0^1 \widehat{p}_i^{+*}(u\lambda, \tau_i) du \right), \\
 &= \int_0^1 \mathbb{E}(\widehat{p}_i^{+*}(u\lambda, \tau_i)) du, \\
 &= \int_0^1 \widehat{p}_i^*(u\lambda) du, \quad (\text{from [?, page 879]}) \\
 &= p_i^*.
 \end{aligned}$$

The pricing scheme is optimal because the average price over periods for a class- i job is the optimal price p_i^* .

Incentive Compatibility From [?], we know that the pricing scheme $\widehat{p}_i^{+*}(\lambda, t)$ is incentive compatible,

$$\widehat{p}_i^{+*}(\lambda^*) + v_i (W_i^q(\lambda^*) + c_i) < \widehat{p}_j^{+*}(\lambda^*) + v_i (W_j^q(\lambda^*) + c_j),$$

for all $j \in \{1, \dots, n\} \setminus \{i\}$.

Hence, due to the non-negativity of $\widehat{p}_i^{+*}(\lambda^*)$, one can assure that our pricing scheme is incentive compatible too by an integration. \blacksquare

In the next sections, we explore two other extensions about: network modeling and dynamic pricing.

5 Network extension

This section and the following are based on [?] in which there are two extensions to the marginal cost model of [?] : the queuing network case and the time-dynamic case. Both extensions use the marginal cost mechanism. We consider a network composed of K nodes indexed by $k \in \{1, 2, \dots, K\}$. Jobs enter in the network through node k with rate Λ_k . Arrivals of class- i jobs constitute a Poisson process with rate Λ^i . Each job has a route within the network denoted by r which is a series of nodes it visits. All jobs are stochastically identical in length and served in FCFS at each node. The processing time at node k is exponentially distributed with rate μ_k . We denote by Q^i the transition matrix

where $Q_{k',k}^i$ is the probability that a class- i job transits from node k' to node k . The route for a class- i job is then a random variable r^i . The destination node of a class- i job is an absorbing state in Q^i . Also, we note \underline{q}^i the vector where q_k^i is the probability that a class- i job arrives to the network through node k . Denoting by W_k the expected waiting time at node k , the expected congestion cost to a class- i job is $vW^i = v\mathbb{E}(\sum_{k \in r^i} W_k)$ where the expectation is for the random route.

If we call λ_k^i the effective arrival rate of class- i jobs to node k , we have the following relation between effective and external rates [?]:

$$\underline{\lambda}^i = \Lambda^i \underline{b}^i, \quad (14)$$

with $\underline{b}^i = \underline{q}^i(I - Q^i)^{-1}$, $\underline{\lambda}^i = (\lambda_1^i, \lambda_2^i, \dots, \lambda_K^i)$, $\underline{q}^i = (q_1^i, q_2^i, \dots, q_K^i)$ and where I is the identity matrix.

If the charge for a class- i job is p^i , the job arrival rate is determined by the demand relationship

$$\int_0^1 (V^i)'(\Lambda^i t) dt = \frac{V^i(\Lambda^i)}{\Lambda_i} = p^i(\Lambda^i) + v \int_0^1 W^i(\underline{\Lambda} t) dt. \quad (15)$$

We then have the following proposition which gives the optimal price with the last demand relationship.

Proposition 2 *The optimal price p^i for a class- i job is*

$$p^i(\Lambda^*) = \sum_{k \in r^i} b_k^i p_k(\Lambda^*), \quad (16)$$

where the price at node k is

$$p_k(\Lambda^*) = v \sum_i (b_k^i \Lambda^{i*}) \int_0^1 t W'_k \left(t \sum_i b_k^i \Lambda^{i*} \right) dt.$$

If we suppose that, $\forall k$, node k is a $M/M/1$ queue with service requirement μ_k , we have

$$p_k(\Lambda^*) = \frac{v \ln \left(\frac{\mu_k - \sum_i (b_k^i \Lambda^{i*})}{\mu_k} \right)}{\sum_i b_k^i \Lambda^{i*}} + \frac{v}{\mu_k - \sum_i b_k^i \Lambda^{i*}}. \quad (17)$$

Proof The social welfare maximization is :

$$\begin{aligned} \max_{\underline{\Lambda}} NV(\underline{\Lambda}) &= \max_{\underline{\Lambda}} \left(\sum_i V^i(\Lambda^i) - \sum_i v \Lambda^i W^i(\underline{\Lambda}) \right), \\ &= \max_{\underline{\Lambda}} \left(\sum_i V^i(\Lambda^i) - \sum_k v \left(\sum_i b_k^i \Lambda^i \right) W_k \left(\sum_i b_k^i \Lambda^i \right) \right), \end{aligned} \quad (18)$$

where we have

$$W^i = \mathbb{E} \left(\sum_{k \in r^i} W_k \right) = \sum_{k \in r^i} b_k^i W_k \quad \text{and} \quad W_k(\lambda_1, \lambda_2, \dots, \lambda_K) = W_k \left(\sum_{i=1}^I \lambda_k^i \right), \quad (19)$$

because b_k^i is the average number of visits to node k by a class- i job. The first order condition for social welfare maximization Equation (18) gives:

$$(V^i)'(\Lambda^{i*}) = v \sum_{k \in r^i} b_k^i W_k \left(\sum_{i=1}^n b_k^i \Lambda^{i*} \right) + \sum_k v \left(\sum_i b_k^i \Lambda^{i*} \right) b_k^i W'_k \left(\sum_i b_k^i \Lambda^{i*} \right).$$

Using the demand relationship (15), we then obtain

$$\begin{aligned} p^i(\Lambda^*) &= \int_0^1 \sum_k v \left(\sum_i b_k^i \Lambda^{i*} t \right) b_k^i W'_k \left(t \sum_i b_k^i \Lambda^{i*} \right) dt, \\ &= \sum_k b_k^i v \left(\sum_i b_k^i \Lambda^{i*} \right) \int_0^1 t W'_k \left(t \sum_i b_k^i \Lambda^{i*} \right) dt, \\ &= \sum_k b_k^i p_k(\Lambda^*). \end{aligned}$$

We have from [?] the expression of the waiting time for a job in a M/M/1 queue with arrival rate λ and service requirement μ fixed

$$W(\lambda) = \frac{\lambda}{\mu - \lambda},$$

so the first derivative is

$$W'(\lambda) = \frac{1}{(\mu - \lambda)^2}.$$

Thus, in our case, we have a M/M/1 queue for each node k with arrival rate $\sum_{i=1}^n \lambda_k^i$ which is equal to $\sum_i (b_k^i \Lambda^i)$ by employing relation (14). Hence, we have

$$W'_k(t \sum_i (b_k^i \Lambda^{i*})) = \frac{1}{(\mu_k - t \sum_i (b_k^i \Lambda^{i*}))^2},$$

and thus we obtain:

$$p_k(\Lambda^*) = \frac{v \ln\left(\frac{\mu_k - \sum_i (b_k^i \Lambda^{i*})}{\mu_k}\right)}{\sum_i b_k^i \Lambda^{i*}} + \frac{v}{\mu_k - \sum_i b_k^i \Lambda^{i*}}.$$

■

In the next step, we consider the time-dynamic pricing scheme, where the rates vary over time.

6 Dynamic pricing with Markovian expectations

This time-dynamic extension is studied just for one class of jobs for simplicity, but it can be extended to multi-class case giving only more complex expressions.

We suppose that the rates are constant during small units of time. The time is then divided into an infinite number of non-overlapping periods. At the end of a given period t , the price in the next period $t + 1$ is computed on the basis of the arrival rate Λ_t . The price is determined by the optimal expression (17). And so we have the following relation

$$p_{t+1}^+ = p(\Lambda_t), \quad (20)$$

with

$$p_k(\Lambda_t) = v \sum_i (b_k^i \Lambda_t) \int_0^1 x W'_k \left(x \sum_i b_k^i \Lambda_t \right) dx. \quad (21)$$

The price over period $t + 1$ depends on the arrival rate for the previous period t . Moreover, arrival rate for the period $t + 1$ is estimated in terms of those at the previous period t . Hence, we have Markovian expectations defined by

$$[\Lambda_{t+1} | \bar{\Lambda}_t] = \Lambda_t, \quad (22)$$

where $[\Lambda_{t+1}|\bar{\Lambda}_t]$ is the arrival rate for the period $t+1$ at the end of the precedent period t . The new arrival rate Λ_{t+1} verifies the following demand relationship

$$\frac{V^i(\Lambda_{t+1}^i)}{\Lambda_{t+1}^i} = \sum_k b_k^i p_k(\Lambda_t) + v \int_0^1 W^i(x[\Lambda_{t+1}|\bar{\Lambda}_t]) dx. \quad (23)$$

We can see, with respect to the demand relationship (8) that the valuation at period $t+1$ depends on the expected rates deduced from those obtained at period t . We can then define a function f as

$$\Lambda_{t+1} = f(\Lambda_t).$$

But, as we must verify the ergodicity inequality $\Lambda_t < 1$, we define for $\epsilon > 0$ and for all periods t :

$$\tilde{f}(\Lambda_t) = \begin{cases} \epsilon & \text{if } f(\Lambda_t) < 0, \\ 1 - \epsilon & \text{if } f(\Lambda_t) > 1 - \epsilon, \\ f(\Lambda_t) & \text{otherwise,} \end{cases}$$

in order to ensure that the rates are not negative or superior of the service rate $\mu = 1$ [?].

Example We consider a M/M/1 queue with jobs value function $V(\Lambda) = 5\Lambda - 5\Lambda^2$, $\mu = 1$, $v = 1$ and $\epsilon = 0.05$. The demand relationship (23) is

$$V(\Lambda_{t+1}) = \Lambda_{t+1} \int_0^1 (\Lambda_t x W'(x\Lambda_t) + W(x\Lambda_t)) dx, \quad (24)$$

with $W(\Lambda) = \frac{\Lambda}{1-\Lambda}$, giving

$$V(\Lambda_{t+1}) = \Lambda_{t+1} W(\Lambda_t) = \Lambda_{t+1} \frac{\Lambda_t}{1 - \Lambda_t},$$

so we obtain the relation

$$\Lambda_{t+1} = \frac{1}{5} \left(5 - \frac{\Lambda_t}{1 - \Lambda_t} \right) = f(\Lambda_t).$$

Figure 1 illustrates the instability of the arrival rate, therefore of the price is this case.

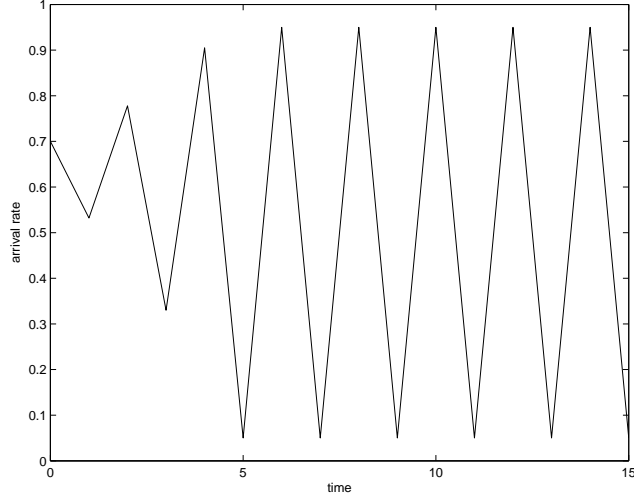


Figure 1: Arrival rate with respect to the time

Theorem 5 *Consider a network reduced to a simple $M/M/1$ queue with single-class users, service requirement $\mu = 1$ and Markovian expectations defined by (20) and (22). Then the system is stable, i.e. arrival rates converge, if and only if*

$$\frac{v}{(1 - \Lambda^*)^2} < \frac{|v(1 - \Lambda^*)\Lambda^*V'(v\frac{\Lambda^*}{1-\Lambda^*}) - (1 - \Lambda^*)^2V(v\frac{\Lambda^*}{1-\Lambda^*})|}{v^2\Lambda^{*2}}, \quad (25)$$

where Λ^* is the solution of the fixed point equation $\Lambda = f(\Lambda)$.

Proof Since the number of nodes is $K = 1$, the average number of visits to node k by a class- i job is $b_k^i = 1$. Using (23), we define the following function

$$\begin{aligned} U(\Lambda_{t+1}) &= \frac{V(\Lambda_{t+1})}{\Lambda_{t+1}} = p(\Lambda_t) + v \int_0^1 W^i(x[\Lambda_{t+1}|\bar{\Lambda}_t])dx, \\ &= p(\Lambda_t) + v \int_0^1 W^i(x\Lambda_t)dx, \\ &= g(\Lambda_t), \end{aligned} \quad (26)$$

so that we have

$$\Lambda_{t+1} = U^{-1}(g(\Lambda_t)) = f(\Lambda_t).$$

A condition for convergence is $|f'(\Lambda^*)| < 1$, with Λ^* solution of the fixed point equation $\Lambda^* = f(\Lambda^*)$. The first derivative of the function f in the optimal rate is

$$\begin{aligned} f'(\Lambda^*) &= (U^{-1} \circ g)'(\Lambda^*), \\ &= (U^{-1})'(g(\Lambda^*))g'(\Lambda^*). \end{aligned}$$

But, using Equations (21) and (26),

$$\begin{aligned} g(\Lambda^*) &= v\Lambda^* \int_0^1 \frac{t}{(1-t\Lambda^*)^2} dt + v\Lambda^* \int_0^1 \frac{t}{1-t\Lambda^*} dt, \\ &= v \frac{\Lambda^*}{1-\Lambda^*}, \end{aligned}$$

giving

$$g'(\Lambda^*) = \frac{v}{(1-\Lambda^*)^2}.$$

We have the relation between U and the value function V :

$$U'(g(\Lambda^*)) = \frac{v(1-\Lambda^*)\Lambda^*V'(v\frac{\Lambda^*}{1-\Lambda^*}) - (1-\Lambda^*)^2V(v\frac{\Lambda^*}{1-\Lambda^*})}{v^2\Lambda^{*2}},$$

since $U'(\Lambda_{t+1}) = \frac{V'(\Lambda_{t+1})}{\Lambda_{t+1}} - \frac{V(\Lambda_{t+1})}{\Lambda_{t+1}^2}$. The condition $|f'(\Lambda^*)| < 1$ becoming $|g'(\Lambda^*)| < |U'(g(\Lambda^*))|$, we obtain the theorem. \blacksquare

This result can easily be generalized to a network with n classes. The stability condition (25) becoming

$$\sum_{l=1}^n \left| \frac{v}{U'(g(\Lambda^{*i}))} \left[\frac{\sum_k b_k^i b_k^l}{\sum_j b_k^j \Lambda^{*j}} \left(\frac{2 - \sum_j b_k^j \Lambda^{*j}}{1 - \sum_j b_k^j \Lambda^{*j}} \right) - 3 \frac{\sum_k b_k^i b_k^l}{(1 - \sum_j b_k^j \Lambda^{*j})^2} \right] \right| < 1.$$

Example (continued) One can check that the optimal rate is $\Lambda^* = 0.64$ by solving the equation $f(\Lambda) = \Lambda$. Then, if we look at the stability condition, and we can see that $|g'(\Lambda^*)| = \frac{1}{(1-0.64)^2} = 7.71$ and $|U'(g(\Lambda^*))| = b = 5$ thus the fixed point condition defined in theorem 5 is not satisfied.

Consider a second example with value function $V(\Lambda) = 5\Lambda - 10\Lambda^2$. We keep all the other variables of the previous example. Here, the optimal rate is $\Lambda^* = 0.4285$. If we look at the stability condition, we obtain $|g'(\Lambda^*)| = \frac{1}{(1-0.42)^2} = 2.97$ and $|U'(g(\Lambda^*))| = 10$ and thus Condition (25) is verified in this case. We obtain, in Figure 2, the convergence of rates with respect to the time.

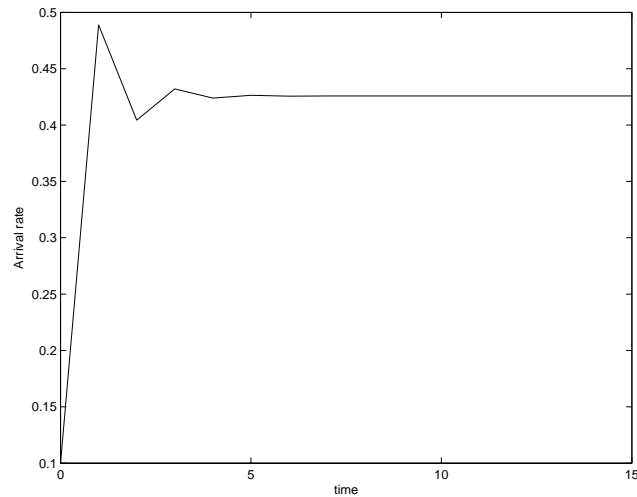


Figure 2: Arrival rate with respect to time

7 Conclusions

This paper studies an optimal priority pricing scheme based on a cost-sharing mechanism. We have modeled communication networks in a multi-class context, as in the marginal cost sharing literature, and then we have optimized the social welfare thanks to the optimal rates and prices. Finally, we have extended the model to the heterogeneous, dynamic and network cases.

As directions for future researches, we can apply our mechanism to different general queuing disciplines or to the case of finite buffers like in [?]. Also we could more specific in terms of Internet traffic classes and requirements.



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irista, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399