On Load Balancing Equilibria in Multiqueue Systems with Multiclass Traffic

Tejas Bodas, D. Manjunath

On Load Balancing Equilibria in Multiqueue Systems with Multiclass Traffic

Tejas Bodas and D Manjunath
IIT Bombay, INDIA

***Abstract*—We consider a queueing system with two non identical FCFS servers together serving two classes of customers. All customers have i.i.d service requirements. One of the queues may charge an admission price, say c. Arrivals are randomly routed to one of the servers and the routing probabilities are determined centrally to optimise a global objective, or from a local mechanism minimising a local—class or individual—objective. Our interest is to analyse the use of c to achieve a target distribution of loads among the servers. We first analyse the structure of the optimal allocation and then consider (1) a system with a dispatcher for each class, (2) a non atomic system, and (3) a system where one of the classes has a dispatcher.**

I. Introduction

We consider multiqueue systems, each FCFS queue with its own server, serving a multiclass population. Customers arrive according to a Poisson process with an intrinsic rate per class. The different classes have the same service requirement but different costs per unit waiting time. The servers have different service rates and may also impose an admission price on each customer joining the queue. Arrivals are randomly routed to a queue and the routing probabilities either minimize a global or a local objective.

Centralised, non competitive, probabilistic allocation of multiclass traffic to multiple M/G/1 servers to minimize the mean waiting time per customer is considered in [1], [2]. FCFS scheduling is analysed in [1] and optimal scheduling is analysed in [2]. In a more distributed system, each class has a dispatcher and Dispatcher i randomly allocates Class i customers to the servers to minimise the waiting cost per Class i customer. This leads to a competition between the dispatchers and is analysed in [3]. Also, [4] considered a single customer class and compared a centralised allocation with individually optimal schemes. Note that none of these systems use an admission price for the queues.

A decentralised system in which each customer randomly chooses a queue to individually optimise its cost is considered in [5]. Here one of the queues charges an admission price to every arriving customer while the second queue does not. For this non atomic system, the equilibrium, as a function of the admission price was characterised in [5]. A system in which customers of one class were routed by a dispatcher while those of a second class made individually optimal decisions was also considered in [5]. Once again, the equilibrium load was analysed. The objective there was to analyse the effect of pricing as a control to achieve a specific load distribution. Some earlier literature on admission price mechanisms include the Paris Metro pricing scheme of [6] and the Tirupati pricing schemes of [7], [8].

In this paper, we continue with the two-queue, two-class system of [5]. One of the queues can charge an admission price. We first consider a centralised system in which the routing probabilities are chosen to optimise a global objective. We characterize this optimal load distribution in Section II. In Section III we analyse the equilibrium loads in a system where the allocation for each class is by a dispatcher. Here one of the queues has an admission price and we obtain the relation between the system parameters for different equilibria. In Section IV we describe a non atomic system where each customer makes an individually optimal decision and characterise the c that will make the equilibrium distribution to be the optimal distribution derived in Section II. Finally, in Section V we analyse the equilibrium structure when one of the classes has a dispatcher while the other class of customers are non atomic.

Before proceeding, we formally describe the model and the notation. Class i customers arrive according to a stationary Poisson process of rates λ_i. Each customer requires a service time that is exponentially distributed with unit mean. A Class i customer has a cost of β_i per unit waiting time; we assume $\beta_1 > \beta_2$. Server of Queue j has a service rate of μ_j; we assume $\mu_1 > \mu_2$. Server of Queue j also charges a per customer admission price c_j. Without loss of generality, $c_1 = c > 0$ and $c_2 = 0$. p_i and q_i denote the fraction of Class i traffic allocated to Server 1 and Server 2 respectively; of course $p_i = 1 - q_i$. γ_j denotes the arrival rate of the total traffic at Queue j; clearly $\gamma_1 = p_1\lambda_1 + p_2\lambda_2$, $\gamma_2 = q_1\lambda_1 + q_2\lambda_2$ and $\gamma_1 + \gamma_2 = \lambda_1 + \lambda_2$. For stability of the queuing system we assume that $\mu_1 + \mu_2 > \lambda_1 + \lambda_2$.

Let the expected waiting time in the Queue j be $D_j(\gamma_j) = \frac{1}{\mu_j - \gamma_j}$ for $j = 1, 2$. The expected total cost (sum of the expected waiting cost and the admission price) per Class i customer, denoted by Δ_i, is

$$\Delta_i = (p_i(c + \beta_i D_1(\gamma_1)) + q_i\beta_i D_2(\gamma_2)),$$

while the overall social cost when $c = 0$ is defined as

$$\begin{aligned}\Delta_s &= = \frac{\lambda_1\Delta_1 + \lambda_2\Delta_2}{\lambda_1 + \lambda_2} \\ &= \frac{(p_1\lambda_1\beta_1 + p_2\lambda_2\beta_2)(D_1(\gamma_1) - D_2(\gamma_2))}{\lambda_1 + \lambda_2} \\ &\quad + \frac{(\lambda_1\beta_1 + \lambda_2\beta_2) D_2(\gamma_2)}{\lambda_1 + \lambda_2} \qquad (1)\end{aligned}$$

Tejas Bodas is funded by the Bharti Centre for Communication at IIT Bombay

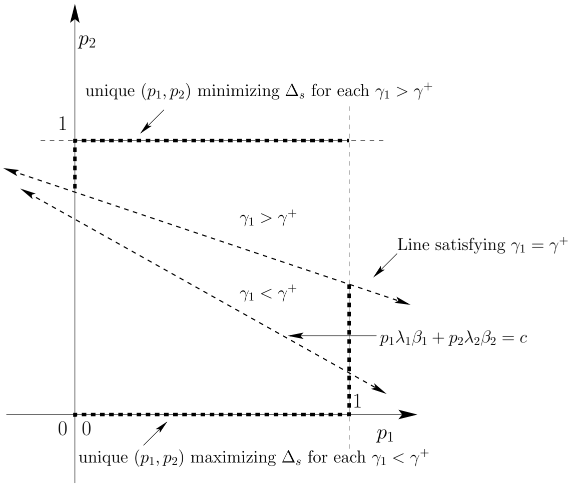
p_2
unique (p_1, p_2) minimizing Δ_s for each $\gamma_1 > \gamma^+$
1
$\gamma_1 > \gamma^+$
Line satisfying $\gamma_1 = \gamma^+$
$\gamma_1 < \gamma^+$
$p_1\lambda_1\beta_1 + p_2\lambda_2\beta_2 = c$
1
0
0
p_1
unique (p_1, p_2) maximizing Δ_s for each $\gamma_1 < \gamma^+$

Fig. 1. Structure of (p_1^*, p_2^*) minimizing Δ_s for arbitrary γ^+

Also define $\Theta = \{\lambda_1, \lambda_2, \mu_1, \mu_2, \beta_1, \beta_2\}$ to denote the set of system parameters.

II. Optimal Allocation in a Centralised Scheme

We consider the centralised routing of customers to the servers to minimize the social cost Δ_s. For this centralised scheme we consider the admission price $c = 0$. Let (p_1^*, p_2^*) be the routing probabilities corresponding to the optimal allocation. For the stability of the queues, we require that $\mu_1 > \gamma_1$ and $\mu_2 > \gamma_2$ (or equivalently, $\gamma_1 > \lambda_1 + \lambda_2 - \mu_2$). Thus the feasible allocation space for (p_1, p_2) is obtained as $\mathscr{S} = \{(p_1, p_2) : \gamma_1 < \mu_1, \gamma_1 > (\lambda_1 + \lambda_2 - \mu_2)\}$; of course, $0 \leq \gamma_1 \leq \lambda_1 + \lambda_2$. We also define $\gamma^+ = \{\gamma_1 : D_1(\gamma_1) = D_2(\gamma_2)\}$. It can be shown that

$$\gamma^+ = \frac{\mu_1 - \mu_2 + \lambda_1 + \lambda_2}{2} \tag{2}$$

While for all feasible $\gamma_1 < \gamma^+$ we have $D_1(\gamma_1) < D_2(\gamma_2)$, for $\gamma_1 > \gamma^+$ we have $D_1(\gamma_1) > D_2(\gamma_2)$. We now have the following properties.

Property 1: For a fixed $\gamma_1 \neq \gamma^+$, there exists a unique (p_1, p_2) pair that minimizes Δ_s at the particular γ_1.

Proof: Consider the case where $0 \leq \gamma^+ \leq \lambda_1 + \lambda_2$; the arguments are similar when this is not the case. For a fixed $\gamma_1 < \gamma^+$, $D_1(\gamma_1)$ and $D_2(\gamma_2)$ are constant and since $D_1(\gamma_1) < D_2(\gamma_2)$, from (1) we see that Δ_s is minimized by the following linear program.

$$\begin{aligned} \arg\max_{p_1, p_2} \quad & p_1\lambda_1\beta_1 + p_2\lambda_2\beta_2 \\ \text{subject to} \quad & p_1\lambda_1 + p_2\lambda_2 = \gamma_1 \end{aligned} \tag{3}$$

In this linear program, if $\beta_1 \neq \beta_2$, then the slope of the objective function and that of the constraint are different. Thus there exists a unique (p_1, p_2) at the boundary of the feasible space which is a solution to (3). For $\gamma_1 > \gamma^+$, the $\arg\max$ in (3) is replaced by $\arg\min$ and the same argument applies.

The above is illustrated in Fig. 1 where the dark dotted lines represent the solution to 3 for different values of the equality constraint γ_1. □

We can now analyses the structure of (p_1^*, p_2^*). Denote the γ_1 corresponding to (p_1^*, p_2^*) as γ_1^*. If Θ is such that $\gamma_1^* \neq \gamma^+$, then the optimal allocation (p_1^*, p_2^*) is at the boundary of the feasible space. This follows from Property 1 and is illustrated in Fig. 1. If Θ is such that $\gamma_1^* = \gamma^+$, then $D_1(\gamma_1) = D_2(\gamma_2)$ and any (p_1, p_2) satisfying $p_1\lambda_1 + p_2\lambda_2 = \gamma^+$ is a valid(p_1^*, p_2^*).

We also remark that if $\gamma_1^* = \gamma^+$, then choosing a (p_1^*, p_2^*) that is not at the boundary would result in a higher variance of the waiting cost than that from the boundary. This is because in the former case, the traffic mix at each of the server will be more heterogeneous than that at the boundary. Thus even when $\gamma_1^* = \gamma^+$, a (p_1^*, p_2^*) at the boundary of the feasible space may be preferred.

Property 2: There are at most four minimizers (p_1^*, p_2^*).

Proof: The boundary of the set of feasible (p_1, p_2) consists of four parts—$B_1 = \{(p_1, p_2) : p_1 = 1, p_2 \in [0, 1]\}$, $B_2 = \{(p_1, p_2) : p_1 \in [0, 1], p_2 = 0\}$, $B_3 = \{(p_1, p_2) : p_1 = 0, p_2 \in [0, 1]\}$, and $B_4 = \{(p_1, p_2) : p_1 \in [0, 1], p_2 = 1\}$. First consider B_1. Since $p_1 = 1$, $\Delta_1 = \frac{\beta_1}{\mu_1 - \lambda_1 - p_2\lambda_2}$. It can be shown that Δ_1 is convex in p_2. Similarly $\Delta_2 = \frac{\beta_2 p_2}{\mu_1 - \lambda_1 - p_2\lambda_2} + \frac{\beta_2 q_2}{\mu_2 - q_2\lambda_2}$ is a sum of two convex functions and hence is convex in p_2. Since Δ_s is a weighted sum of two convex functions in p_2, Δ_s is also convex in p_2 with a unique minimizer. Arguing along similar lines, Δ_s is convex in p_2 in set B_3, while it is convex in p_1 in sets B_2 and B_4. This completes the proof. □

We can make the following remarks from the above properties.

- Extensive numerical investigation indicates that $\gamma^* < \gamma^+$ implying $D_1(\gamma_1) < D_2(\gamma_2)$. Further, we also see that (p_1^*, p_2^*) is unique. However we do not have a formal proof and we leave that as an open problem for future work.
- At the optimal allocation, if $\gamma_1^* < \gamma^+$ then $D_1(\gamma_1) < D_2(\gamma_2)$ and either $p_1^* = 1$ and $p_2^* \in [0, 1]$ or $p_1^* \in (0, 1]$ and $p_2^* = 0$.
- If $\gamma_1^* > \gamma^+$ then $D_1(\gamma_1) > D_2(\gamma_2)$ and hence $p_2^* = 1$ and $p_1^* \in [0, 1]$ or $p_2^* \in (0, 1]$ and $p_1^* = 0$.
- From the preceding two remarks, we see that Class 1 traffic 'has priority over Class 2' in the choice of server. At (p_1^*, p_2^*) Class 1 traffic utilises the server with a lower expected cost while the Class 2 traffic is forced to use a server with a higher expected waiting time.

III. Equilibrium Loads with Two Dispatchers

We now consider a decentralised system in which each traffic class has a dispatcher. Queue 1 charges an admission price $c > 0$ per customer and Queue 2 does not. Dispatcher i chooses a strategy p_i to minimize Δ_i, the expected total cost of a Class i customer. The optimal strategy for each dispatcher depends on the strategy of the other dispatcher. A Nash equilibrium strategy pair for the dispatchers is then defined as the allocation $(\hat{p}_1, \hat{p}_2)$ from which there is no incentive for either dispatcher to change its strategy. From Theorem 2.1 in [9], we see that Δ_1 and Δ_2 satisfy the conditions for the

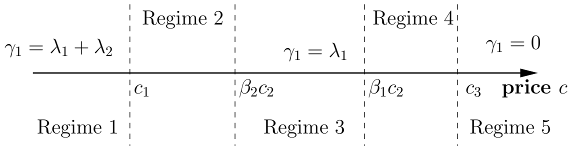

Fig. 2. The operating regimes as c is increased from 0.

existence of a Nash equilibrium. The equilibrium allocation, $(\hat{p}_1, \hat{p}_2)$, can be obtained as the solution to the following set of equations derived from the first order condition on Δ_1 and Δ_2.

$$c + \frac{\mu_1 - p_2\lambda_2}{(\mu_1 - \gamma_1)^2} - \frac{\mu_2 - q_2\lambda_2}{(\mu_2 - \gamma_2)^2} = 0.$$
$$c + \frac{\mu_1 - p_1\lambda_1}{(\mu_1 - \gamma_1)^2} - \frac{\mu_2 - q_1\lambda_1}{(\mu_2 - \gamma_2)^2} = 0$$

An explicit solution to the above set of equations is messy. Rather than elaborate on that, we now find conditions on the parameter set Θ for specific types of equilibria.

First consider the equilibrium $\hat{p}_1 = \hat{p}_2 = 1$. For this we require that at $p_1 = p_2 = 1$, $\frac{\partial \Delta_1}{\partial p_1} \leq 0$ and $\frac{\partial \Delta_2}{\partial p_2} \leq 0$. Thus

$$c + \frac{\mu_1 - \lambda_2}{(\mu_1 - \lambda_1 - \lambda_2)^2} - \frac{1}{\mu_2} \leq 0.$$
$$c + \frac{\mu_1 - \lambda_1}{(\mu_1 - \lambda_1 - \lambda_2)^2} - \frac{1}{\mu_2} \leq 0.$$

This implies that if the queue stability condition $\mu_1 > (\lambda_1 + \lambda_2)$ is satisfied, $(\hat{p}_1 = 1, \hat{p}_2 = 1)$ if

$$\mu_2 \leq \frac{(\mu_1 - \lambda_1 - \lambda_2)^2}{c(\mu_1 - \lambda_1 - \lambda_2)^2 + (\mu_1 - \max(\lambda_1, \lambda_2))} \qquad (4)$$

Similarly for $\hat{p}_1 = \hat{p}_2 = 0$ we require

$$\mu_1 \leq \frac{(\mu_2 - \lambda_1 - \lambda_2)^2}{c(\mu_1 - \lambda_1 - \lambda_2)^2 + (\mu_2 - \max(\lambda_1, \lambda_2))}$$

which implies $\mu_1 < \mu_2$. Recall that we have assumed that $\mu_1 > \mu_2$; hence this equilibrium is not possible.

For $\hat{p}_1 = 1, 0 < \hat{p}_2 < 1$ to be a valid equilibrium the necessary condition on the parameters are obtained from the conditions $\frac{\partial \Delta_1}{\partial p_1} \leq 0$ and $\frac{\partial \Delta_2}{\partial p_2} = 0$ at $(p_1 = 1, 0 < p_2 < 1)$. Thus we need p_2 that solves

$$\frac{\mu_1 - \lambda_1}{(\mu_1 - \lambda_1 - p_2\lambda_2)^2} \quad = \quad \frac{\mu_2 - q_1\lambda_1}{(\mu_2 - q_2\lambda_2)^2}$$

and satisfies

$$c + \frac{\mu_1 - p_2\lambda_2}{(\mu_1 - \lambda_1 - p_2\lambda_2)^2} \quad - \quad \frac{\mu_2 - q_2\lambda_2}{(\mu_2 - q_2\lambda_2)^2} \leq 0.$$

The necessary conditions on system parameters for other equilibrium can be obtained similarly. Numerical results to illustrate the equilibria will be provided in the final version.

IV. Equilibrium Loads in a Non Atomic System

The equilibrium load distribution of the non atomic system with two classes was studied in [5]. Recall that in such a system, each customer makes an individually optimal queue-join decision by joining a queue that minimizes its expected total cost. Of course the total cost is the sum of the admission price and the waiting cost. In [5], for a fixed Θ, the equilibrium traffic distribution, $(\hat{p}_1, \hat{p}_2)$, as a function of c was characterized. It was shown that the traffic at equilibrium is in one of the following five regimes. (1) Regime 1 for which $\hat{p}_1 = \hat{p}_2 = 1$, (2) Regime 2 for which $\hat{p}_1 = 1$ and $0 < \hat{p}_2 < 1$, (3) Regime 3 for which $\hat{p}_1 = 1$ and $\hat{p}_2 = 0$, (4) Regime 4 for which $0 < \hat{p}_1 < 1$ and $\hat{p}_2 = 0$, and (5) Regime 5 for which $\hat{p}_1 = 0$ and $\hat{p}_2 = 0$. This is illustrated in Fig. 2. Note that in Fig. 2, c_1, c_2 and c_3 are a function of Θ, the system parameters. Since we assume that $c > 0$, the different regimes are feasible if $0 < c_1 < c_2 < c_3$. For example, it is shown in [5] that $c_1 > 0$ requires $\mu_1 > \mu_2 + \lambda_1 + \lambda_2$. Also, note that the five regimes of the non atomic model correspond to the boundary sets B_1 and B_2 defined in Section II. We thus have the following.

Lemma 1: If $\mu_1 > \mu_2 + \lambda_1 + \lambda_2$, then there exists an admission price c with a corresponding non atomic traffic equilibrium $(\hat{p}_1, \hat{p}_2)$ which is also the minimizer of social cost Δ_s i.e. $(\hat{p}_1, \hat{p}_2) = (p_1^*, p_2^*)$.

Proof: Now as $\mu_1 > \mu_2 + \lambda_1 + \lambda_2$, from Eq. 2 it is clear that $\gamma^+ > \lambda_1 + \lambda_2$. Refer Figure 1. As $\gamma_1^* < \gamma^+$, (p_1^*, p_2^*) lies in either B_1 or B_2. As Regimes 1 through 5 correspond to regions on the boundaries B_1 and B_2, we can charge an admission price c to customers joining Server 1 such that the non atomic equilibrium distribution will be (p_1^*, p_2^*) with the corresponding $\gamma_1 = \gamma_1^*$. □

Lemma 1 says that if Θ is such that $\gamma_1^* < \gamma^+$, then a (p_1^*, p_2^*) minimizing the social cost can be achieved in the non atomic system by charging an appropriate admission price c as suggested by Fig. 2. Our numerical results suggest that if $\mu_1 > \mu_2$ and $\beta_1 > \beta_2$, then there exists a unique $\gamma_1^* \leq \gamma^+$ and hence in the non atomic system, the social cost Δ_s can be minimized by charging an appropriate admission price to achieve γ_1^*. At this time, the proof is an open problem.

Note that if $c = 0$, then the equilibrium $(\hat{p}_1, \hat{p}_2)$ is not unique. In fact any (p_1, p_2) that satisfies $p_1\lambda_1 + p_2\lambda_2 = \gamma^+$ is a valid non atomic equilibrium.

V. A Dispatcher for Only One of the Classes

We now consider a model where Class 1 traffic has an associated dispatcher while Class 2 traffic is non atomic. First consider the case of $c = 0$. A Class 2 arrival makes an individually optimal join decision, but the dispatcher chooses a strategy p_1 that minimizes the cost Δ_1 of the Class 1 customer. Clearly, Δ_1 is a function of the traffic distribution of Class 2 traffic; hence for any p_1 chosen by the dispatcher, there is a corresponding equilibrium $\hat{p}_2$ for Class 2 traffic. Once again we have a competitive situation and we are interested in the equilibrium pair $(\hat{p}_1, \hat{p}_2)$.

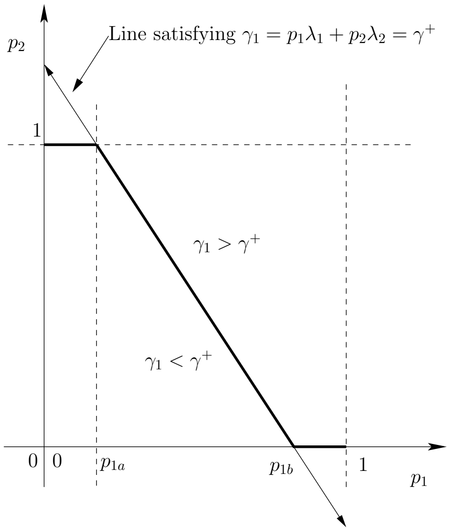

Fig. 3. Feasible $(\hat{p}_1, \hat{p}_2)$ for a particular parameter set Θ. For certain Θ it may be true that $p_{1a} \notin [0,1], p_{1b}$ and $\notin [0,1]$. In that case $\hat{\gamma}_1 = \gamma^+$ is the only feasible region.

Let γ_1 corresponding to $(\hat{p}_1, \hat{p}_2)$ be denoted by $\hat{\gamma}_1 = \hat{p}_1\lambda_1 + \hat{p}_2\lambda_2$. For all $\hat{\gamma}_1 < \gamma^+$, we see that $D_1(\gamma_1) < D_2(\gamma_2)$. This requires that $\hat{p}_2 = 1$. Similarly for $\hat{\gamma}_1 > \gamma^+$, as $D_1(\gamma_1) > D_2(\gamma_2)$ we have $\hat{p}_2 = 0$. For $\hat{\gamma}_1 = \gamma^+$ we have $\hat{p}_2 \in [0,1]$ and $\hat{p}_1\lambda_1 + \hat{p}_2\lambda_2 = \gamma^+$.

Note that the existence of $p_{1a} \in [0,1]$ and $p_{1b} \in [0,1]$ as shown in Fig. 3 may not hold for some Θ. For $\hat{p}_1 < p_{1a}$, $D_1(\gamma_1) < D_2(\gamma_2)$ and hence $\hat{p}_2 = 1$. Similarly For $\hat{p}_1 > p_{1b}$, $D_1(\gamma_1) > D_2(\gamma_2)$ and hence $\hat{p}_2 = 0$. The dark line of Figure 3 is the feasible region for $(\hat{p}_1, \hat{p}_2)$. In [5], we have considered the 1-Dispatcher model with Server 1 charging an admission price c per customer. We characterize the feasible $(\hat{p}_1, \hat{p}_2)$ for different values of c charged by Server 1. For a particular Θ and c, the method to obtain the equilibrium is given in [5]. We now provide a numerical example comparing the non atomic model with this 1-Dispatcher model.

Example: Let $\lambda_1 = \lambda_2 = 2$, $\beta_1 = 2$, $\beta_2 = 1$, $\mu_1 = 12$, and $\mu_2 = 5$. For this Θ, we have $c_1 = 0.0750$, $c_2 = 0.233$ and $c_3 = 1.833$. Fig. 4 shows Δ_s, the the system cost, for the two models for different values of the admission price c. Observe that for lower values of c the system cost in the two models is the same but as the admission price c is increased, the one dispatcher system has a lower per customer cost as compared to the non atomic system.

Fig. 5 shows the equilibrium $\hat{p}_1$ in the two models. Clearly as c is increased, the expected cost per Class 1 customer is smaller with the dispatcher than in the non atomic model.

Finally, in Fig. 6 we see that although the system equilibrium $\hat{p}_2$ in both the models is the same, for higher admission price the per Class 2 customer expected cost also decreases due to the dispatcher for Class 1.

VI. Summary

We began by analysing a centralised allocation of multiclass traffic to heterogeneous servers to minimise the system cost. We analysed the structure of optimal routing probabilities (p_1^*, p_2^*). Decentralisation by having dispatchers for each class yields a competitive system for which we have analysed the equilibrium allocations. A further decentralisation gives us the non atomic system in which each arrival makes its own routing decision. This system was analysed in some detail in [5] and here we described how price could be used to make the equilibrium allocation to be socially optimal. We briefly considered a system where only one of class has a dispatcher. Thus, we have investigated, in some detail, the use of differential admission pricing as a means of decentralised control of load distribution.

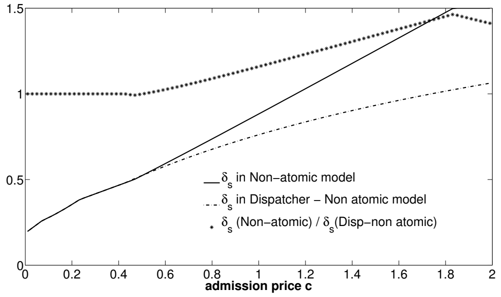

Fig. 4. Comparison of the per customer expected total cost in the two models.

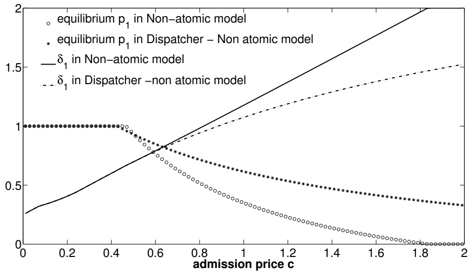

Fig. 5. Comparison of p_1 at system equilibrium for the two models

References

[1] S. C. Borst, "Optimal probabilistic allocation of customer types to servers," in *Proceedings of ACM SIGMETRICS*, September 1995, pp. 116–125.

[2] J. Sethuraman and M. Squillante, "Optimal stochastic scheduling in multiclass parallel queues," in *Proceedings of ACM SIGMETRICS*, May 1999, pp. 93–102.

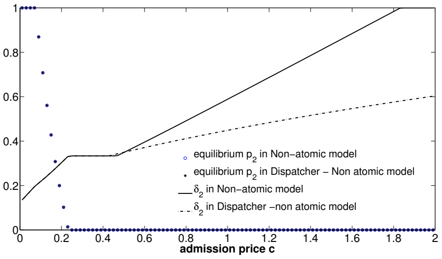

Fig. 6. Comparison of p_2 at system equilibrium in the two models.

[3] U. Ayesta, O. Brun, and B. J. Prabhu, "Price of anarchy in non-cooperative load balancing," in *Proceedings of the IEEE INFOCOM*, 2010, pp. 436–440.

[4] C. H. Bell and S. Stidham, "Individual versus social optimization in the allocation of customers to alternative servers," *Management Science*, vol. 29, pp. 831–839, 1983.

[5] T. Bodas, A. Ganesh, and D. Manjunath, "Load balancing and routing games with admission price," in *Proceedings of the IEEE Conference on Decision and Control (CDC)*, 2011.

[6] A. Odlyzko, "Paris Metro pricing for the internet," in *Proceedings of the 1st ACM Conference on Electronic Commerce*, 1999, pp. 140–147.

[7] P. Dube, V.S. Borkar, and D. Manjunath, "Differential join prices for parallel queues: social optimality, dynamic pricing algorithms and application to internet pricing," in *Proceedings of IEEE INFOCOM*, 2002, pp. 276–283.

[8] V. S. Borkar and D. Manjunath, "Charge-based control of Diffserv-like queues," *Automatica*, vol. 40, pp. 2043–2057, 2004.

[9] A. Orda, R. Rom, and N. Shimkin, "Competitive routing in multi-user communication networks," *IEEE/ACM Transactions on Networking*, pp. 510–521, 1993.