Strategic Routing in Heterogeneous Discriminatory Processor Sharing Queues

Dieter Fiems[1] and Balakrishna J. Prabhu[2]

[1] Ghent University, Departement TELIN, Gent, Belgium
dieter.fiems@UGent.be
[2] LAAS CNRS, University of Toulouse, CNRS, Toulouse, France
balakrishna.prabhu@laas.fr

Abstract. We consider strategic routing for a two-class discriminatory processor queue with an additional cost for joining the premium class. We show that, depending on the specific parameters of the system, various equilibria can coexist, including equilibria where the queueing system is not ergodic for the equilibrium traffic split. We also investigate how the server can select the priority of the classes and the fees charged from the customers in order to maximise its revenue.

1 Introduction

By introducing service differentiation, customers are offered a choice between different service levels at distinct price points. Customers then make a trade-off between the cost of the different service levels and the benefits they bring, for example, in terms of reductions in sojourn times. This trade-off determines the willingness of customers to pay for these services, and therefore also the revenue of the service provider. Queueing theory provides various options to achieve service differentiation, including preemptive and non-preemptive priority disciplines [1], parallelised services and discriminatory processor sharing [2, 3].

In this paper, we explore strategic routing within the context of coupled-resources in which the capacity available at one resource depends upon the demand at all the resources in the network. In particular, we consider two resources coupled by the discriminatory processor-sharing (DPS) mechanism [4]. Processor-sharing queueing systems, including both discriminatory and generalised processor sharing variants, serve as useful models for resource sharing in networking and computer systems. Notably, processor-sharing queues have been effectively utilised to analyse flow-level resource sharing on the Internet, as demonstrated in studies by Massoulié and Roberts [5–7]. The body of research on processor-sharing queueing systems is extensive. For an in-depth review of the theory and its applications, we recommend the surveys by Yashkov and Yashkova [8] on processor sharing, and by Altman *et al.* [9] on discriminatory processor sharing. In the following discussion, we will specifically concentrate on the game-theoretic analysis of processor-sharing queueing systems. Note that the uncoupled-resource model is a special case of coupled-resource one, and has a larger body of literature due to its ease of analysis.

The analysis of routing games with uncoupled-resources started with the seminal work of Orda *et al.* [10] who gave conditions for the existence and uniqueness of the Nash equilibrium. Since then, a large body of literature has appeared on different variants of this game. Altman and Shimkin [11] examine an observable processor-sharing game in which customers decide whether to join the processor-sharing queue after observing the number of customers already present. A similar model is studied by Ben-Shahar *et al.* [12] under the assumption that customers have heterogeneous preferences. Heterogeneity means that one customer may join after observing a certain number of customers while another customer may not join observing the same number of customers. In a concert queueing game, customers choose their arrival time to minimise a cost that depends on the arrival times of other customers. In [13], customers select their arrival times in a processor-sharing queue, where the cost is influenced by both the sojourn time and the deviation from their preferred departure time. Processor-sharing games are also analysed as models for distributed non-cooperative load balancing. In these load-balancing scenarios, customers either choose a server from a set of servers [14] or are directed to a server by a limited number of dispatchers [14, 15]. Finally, a time- and load-dependent fluid processor-sharing queue is used to model the use of park-and-ride systems during rush hour [16].

In comparison, the coupled-resource game has received less attention. Hayel and Tuffin [2] study a DPS queueing game where the utility of joining different classes is a non-negative, decreasing function of the sojourn time. They demonstrate the existence of a unique Wardrop equilibrium for DPS using these utility functions. Fiems [3] investigates revenue management for DPS queues, where the utility decreases linearly with sojourn times. The study shows that a DPS implementation is preferred over a strict priority discipline if customers are allowed to balk. If balking is not allowed, the optimal DPS discipline degenerates to a strict priority discipline, even if customers have heterogeneous evaluations of sojourn times. The work closest to the model of this paper is the one in [17] (see Section 3 in there) in which customers choose the probability of going to the higher priority class in exchange of a payment. It is shown that the equilibrium is either a pure equilibria or a mixed equilibrium depending upon the cost. The former correspond to all the customers going to one of the two classes while, in the latter, customers split non-trivially over the two classes. The model and the results of [17] assume equal service rates for the two classes.

Contributions: We investigate a Markovian DPS queue where both classes have different service rates, and with a fixed cost for joining the premium class. The players want to minimise their mean sojourn time and may want to pay a cost to get better service. The fact that the service rates are heterogeneous gives rise to types of equilibria that are different from the ones in [17]. We show that given the parameters of the DPS queue, multiple equilibia can coexist, including multiple "stable" equilibria (in the sense that customers converge to these equilibria if they start from a nearby strategy and selfishly minimise their cost). Further, there may even be a non-ergodic equilibrium (there is an incentive to converge to a traffic split where the queueing system is not ergodic). At this

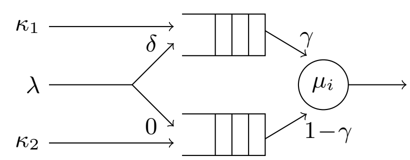


Fig. 1: Representation of the DPS queueing model

non-ergodic equilibrium the queue is unstable and mean sojourn time of both the classes goes to infinity. This type of equilibrium appears to be specific to the coupled-resource model, and in particular to ones with different service rates. The closest result in the uncoupled-resource model that has a similar flavour is the one in [15] in which the Price of Anarchy goes to infinity when the cost of a resource goes to zero. However, for finite costs (as in our paper), instability does not seem to happen in the uncoupled-resource model. Finally, we restrict the parameter space of the DPS model to typical scenarios where the premium tier offers the faster service as well higher DPS weight factors, and discuss how the service provider can optimise its revenue.

Overview: The remainder of this paper is organised as follows. In the next section, we introduce the modelling assumptions and notation, and recall the formulas for the sojourn times in a DPS queueing system. We then discuss the existence of multiple equilibria by means of numerical examples in Section 3. Finally we draw conclusions in Section 4.

2 DPS queueing model

We consider a discriminatory processor-sharing queueing system with two classes as depicted in Figure 1. Non-strategic class 1 and class 2 customers arrive at the queues in accordance with Poisson processes with rates κ_1 and κ_2, respectively. In addition, strategic customers arrive at the queue in accordance with a Poisson process with rate λ and select a class without observing the state of the queueing system. Let λ_i denote the arrival rate of strategic customers that opt for class $i \in \{1,2\}$, with $\lambda_1 + \lambda_2 = \lambda$. The class 1 and class 2 service times constitute a sequence of independent and identically exponentially distributed random variables with rates μ_1 and μ_2, respectively. Without loss of generality, we fix the DPS weight of class 1 and 2 to γ and $1-\gamma$, respectively for $\gamma \in [0,1]$. For $\gamma = 0.5$ there is no differentiation between the classes, while for $\gamma = 1$ ($\gamma = 0$) the DPS discipline degenerates to a strict preemptive priority discipline for class 1 (class 2). This Markovian DPS queueing system admits a stationary solution provided that the load does not exceed the service capacity,

$$\frac{1}{\mu_1}(\kappa_1 + \lambda_1) + \frac{1}{\mu_2}(\kappa_2 + \lambda_2) < 1\,.$$

If this is the case, the expected sojourn times $\bar{T}_1$ and $\bar{T}_2$ of class 1 and 2 equal [9],

$$\overline{T}_1 = \frac{1}{\mu_1(1-\rho)}\left(1 + \frac{\mu_1\rho_2(1-2\gamma)}{\mu_1\gamma(1-\rho_1)+\mu_2(1-\gamma)(1-\rho_2)}\right), \tag{1}$$

$$\overline{T}_2 = \frac{1}{\mu_2(1-\rho)}\left(1 - \frac{\mu_2\rho_1(1-2\gamma)}{\mu_1\gamma(1-\rho_1)+\mu_2(1-\gamma)(1-\rho_2)}\right), \tag{2}$$

with $\rho_i = \mu_i^{-1}(\lambda_i + \kappa_i)$ and $\rho = \rho_1 + \rho_2$.

We now study the fractions of strategic customers that opt for class 1 and class 2, respectively. To this end, we impose that there exist fractions λ_1 and λ_2 with $\lambda_1 + \lambda_2 = \lambda$ such that the queueing process is stationary ergodic. This is the case provided that

$$\frac{\kappa_1}{\mu_1} + \frac{\kappa_2}{\mu_2} + \frac{\lambda}{\max(\mu_1,\mu_2)} < 1\,.$$

The rates λ_1 and λ_2 are now chosen such that none of the customers in these flows has an incentive to change to the other flow. Customers prefer class 1 over class 2 if $\overline{T}_1 + \delta < \overline{T}_2$. Here δ represents a cost for accessing class 1. In other words, we study the rational split of the flows in Wardrop equilibrium. This notion is expressed mathematically as follows,

$$\begin{cases} \overline{T}_1 + \delta \geq \overline{T}_2 & \text{if } \lambda_2 > 0\,, \\ \overline{T}_1 + \delta \leq \overline{T}_2 & \text{if } \lambda_1 > 0\,. \end{cases} \tag{3}$$

Ergodicity: To simplify notation, we can rewrite the expected sojourn times as follows,

$$\overline{T}_1 = \frac{\beta_0 - (\mu_1-\mu_2)(2\gamma-1)\lambda_1}{(\beta_2 - (\mu_1-\mu_2)\lambda_1)\,(\beta_3 - (2\gamma-1)\lambda_1)}\,, \tag{4}$$

$$\overline{T}_2 = \frac{\beta_1 + (\mu_1-\mu_2)(2\gamma-1)\lambda_1}{(\beta_2 - (\mu_1-\mu_2)\lambda_1)\,(\beta_3 - (2\gamma-1)\lambda_1)} \tag{5}$$

with

$$\begin{aligned}
\beta_0 &= \gamma(\kappa_1\mu_2 + 2\kappa_2\mu_1 - \kappa_2\mu_2 + 2\lambda\mu_1 - \lambda\mu_2 - \mu_1\mu_2 + \mu_2^2) \\
&\qquad - \mu_1\kappa_2 + \kappa_2\mu_2 - \mu_1\lambda + \lambda\mu_2 - \mu_2^2\,, \\
\beta_1 &= \gamma(\kappa_1\mu_1 - 2\kappa_1\mu_2 - \kappa_2\mu_1 - \lambda\mu_1 - \mu_1^2 + \mu_1\mu_2) \\
&\qquad + \mu_2\kappa_1 + \mu_1\kappa_2 + \mu_1\lambda - \mu_1\mu_2, \\
\beta_2 &= \mu_2\kappa_1 + \mu_1\kappa_2 + \mu_1\lambda - \mu_1\mu_2, \\
\beta_3 &= \gamma(\kappa_2 - \kappa_1 + \lambda + \mu_1 - \mu_2) - \kappa_2 - \lambda + \mu_2\,.
\end{aligned}$$

With the notation above, we can specify the range Λ of λ_1 for which the queueing process is stationary ergodic,

$$\Lambda = \begin{cases} (\beta_2/(\mu_1-\mu_2), \lambda] \cap [0,\lambda]\,, & \text{for } \mu_1 > \mu_2\,, \\ [0,\lambda]\,, & \text{for } \mu_1 = \mu_2\,, \\ [0, \beta_2/(\mu_1-\mu_2)) \cap [0,\lambda]\,, & \text{for } \mu_1 < \mu_2\,. \end{cases}$$

Note that for $\beta_2/(\mu_1 - \mu_2) = 0$ or $\beta_2/(\mu_1 - \mu_2) = \lambda$, the queueing process is not ergodic for $\lambda_1 = \beta_2/(\mu_1 - \mu_2)$. However, the queueing process is ergodic for $\lambda_1 = 0$ if $\beta_2/(\mu_1 - \mu_2) < 0$. A similar remark applies to the case $\lambda_1 = \lambda$.

In the remainder, we exclude the symmetric case $\gamma = 0.5$ and $\mu_1 = \mu_2$. In this case there is no service differentiation. We then have that all $\lambda_1 \in \Lambda$ are Wardrop equilibria for $\delta = 0$, while $\lambda_1 = 0$ is an equilibrium for $\delta > 0$, and $\lambda_1 = \lambda$ is an equilibrium for $\delta < 0$.

Mixed equilibrium: We now focus on the existence of a mixed equilibrium. A mixed equilibrium lies either in the interior of Λ, or at the non-zero boundary of Λ in case this boundary is not equal to λ. We refer to these types of equilibria as ergodic and non-ergodic equilibria. We discuss ergodic equilibria first.

Solving $\overline{T}_1 + \delta = \overline{T}_2$ leads to the following quadratic equation,

$$\begin{aligned}\delta(2\gamma - 1)(\mu_1 - \mu_2)\lambda_1^2 - ((2\gamma - 1)\delta\beta_2 + (\mu_1 - \mu_2)\,(\delta\beta_3 + 4\gamma - 2))\lambda_1 \\ + \delta\beta_2\beta_3 + \beta_0 - \beta_1 = 0\,.\end{aligned}$$

In the absence of an additional cost ($\delta = 0$), for equal service time ($\mu_1 = \mu_2$) or in absence of service differentiation ($\gamma = \frac{1}{2}$), the quadratic equation simplifies to a linear equation. In this case, we have a single ergodic equilibrium at most.

In general, the quadratic equation may however possess two distinct solutions or possess no real-valued solutions in the interior of Λ. Not every solution is stable though, in the sense that if one slightly deviates from the equilibrium, there is no incentive to drift back towards the equilibrium. A mixed equilibrium is stable provided that

$$\frac{\partial \overline{T}_1}{\partial \lambda_1} > \frac{\partial \overline{T}_2}{\partial \lambda_1}\,. \tag{6}$$

If this inequality holds, an increase of λ_1 translates in longer sojourn times for the first queue, and hence there is an incentive to join the second queue and drift towards the equilibrium.

The condition for stable equilibria simplifies considerably for specific instances. For $\gamma = 0.5$ and $\mu_1 \neq \mu_2$, some elementary calculations show that

$$\frac{\partial \overline{T}_1}{\partial \lambda_1} \Big/ \frac{\partial \overline{T}_2}{\partial \lambda_1} = \frac{\mu_2}{\mu_1}\,.$$

Hence, any equilibrium with $\overline{T}_1 + \delta = \overline{T}_2$ is stable for $\mu_1 < \mu_2$ and unstable for $\mu_1 > \mu_2$. Moreover, for any γ and for $\mu_1 = \mu_2$ we have by means on some elementary calculations,

$$\frac{\partial \overline{T}_1}{\partial \lambda_1} \Big/ \frac{\partial \overline{T}_2}{\partial \lambda_1} = \frac{\overline{T}_1}{\overline{T}_2}\,.$$

In the equilibrium, we have $\overline{T}_1 + \delta = \overline{T}_2$. Hence, positive δ implies $\overline{T}_1 < \overline{T}_2$ and the equilibrium is not stable. Analogously, for negative δ, we have a stable equilibrium.

Pure equilibria: First consider the case $\mu_1 > \mu_2$. We have pure equilibrium for $\lambda_1 = \lambda$ provided that $\overline{T}_1 + \delta \leq \overline{T}_2$. If $\beta_2/(\mu_1 - \mu_2) < 0$, we have a pure equilibrium at $\lambda_1 = 0$ provided that $\overline{T}_1 + \delta \geq \overline{T}_2$. Analogously, for $\mu_1 < \mu_2$, we have pure equilibrium for $\lambda_1 = 0$ provided that $\overline{T}_1 + \delta \geq \overline{T}_2$. If $\beta_2/(\mu_1 - \mu_2) > \lambda$, we have a pure equilibrium at $\lambda_1 = \lambda$ provided that $\overline{T}_1 + \delta \leq \overline{T}_2$. Finally, for $\mu_1 = \mu_2$, the queues are stable for $\lambda_1 = 0$ and $\lambda_1 = \lambda$, and we have two pure equilibria.

A pure equilibrium is stable provided we do not have equality at the equilibrium. If the latter is the case, stability again requires that the customers have an incentive to drift towards the equilibrium, that is, the equilibrium is stable provided (6) holds.

Non-ergodic equilibria: It is possible that customers have an incentive to drift towards a traffic mix such that the queueing process becomes non-ergodic. Now assume $\mu_1 \neq \mu_2$ and $\beta_2/(\mu_1 - \mu_2) \in [0, \lambda]$. Comparing the sojourn times for $\lambda_1 = \beta_2/(\mu_1 - \mu_2)$, we find

$$\frac{\overline{T}_1}{\overline{T}_2} = \frac{\beta_0 - (\mu_1 - \mu_2)(2\gamma - 1)\lambda_1}{\beta_1 + (\mu_1 - \mu_2)(2\gamma - 1)\lambda_1},$$

which simplifies to the surprisingly simple expression,

$$\lim_{\lambda_1 \to \beta_2/(\mu_1 - \mu_2)} \frac{\overline{T}_1 + \delta}{\overline{T}_2} = \frac{(1-\gamma)\mu_2}{\gamma\mu_1}.$$

Note that the limiting value does not depend on δ as both $\overline{T}_1$ and $\overline{T}_2$ go to ∞ while δ remains finite. For $\mu_1 > \mu_2$, there is an incentive to drift towards the non-ergodic traffic mix if the right-hand side exceeds 1. Similarly, for $\mu_1 < \mu_2$, there is an incentive to drift towards the non-ergodic mix if the right-hand side is smaller than 1. Both non-ergodic equilibria are stable. If the right-hand side equals 1, we again have an equilibrium, and stability follows from evaluating

$$\frac{\partial \overline{T}_1}{\partial \lambda_1} \Big/ \frac{\partial \overline{T}_2}{\partial \lambda_1}.$$

As one would most likely want to avoid non-ergodic equilibria, the former conditions easily translate into conditions for the DPS parameter γ. For $\mu_1 > \mu_2$, we impose $\gamma > \mu_2/(\mu_1 + \mu_2)$, while for $\mu_1 < \mu_2$, we need $\gamma < \mu_2/(\mu_1 + \mu_2)$. In words, these conditions limit prioritising the slower server.

Summary: From the discussion above, we may have up to two equilibria in the interior of Λ as well as up to two equilibria at the boundary. As both $\overline{T}_1$ and $\overline{T}_2$ are continuous functions of λ_1, we easily see that stable and unstable equilibria alternate when ordered (accounting for the multiplicity of the solutions of the quadratic equation). Moreover, by listing all possible combinations of pure, mixed and non-ergodic equilibria, we find that the total number of equilibria is at most three.

3 Discussion

From the results above, we have identified multiple different types of equilibrium scenarios. Figure 2 illustrates the different possibilities. For figure 2(a), we set

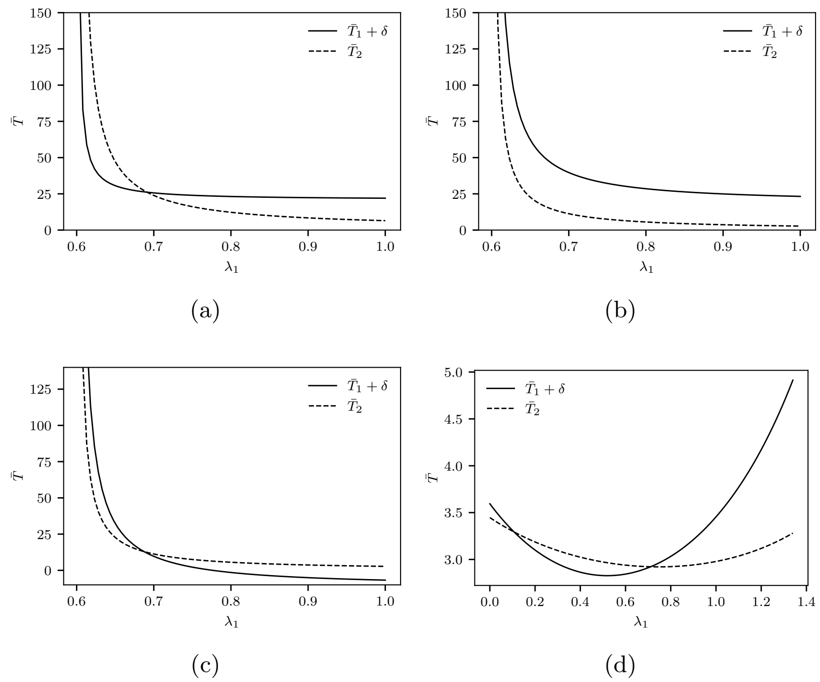
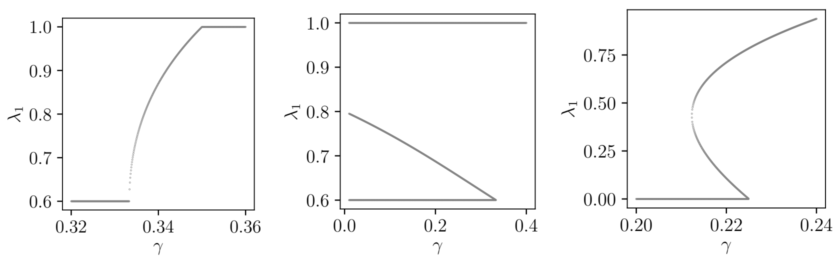


Fig. 2: Different types of equilibria in the DPS game with additional cost

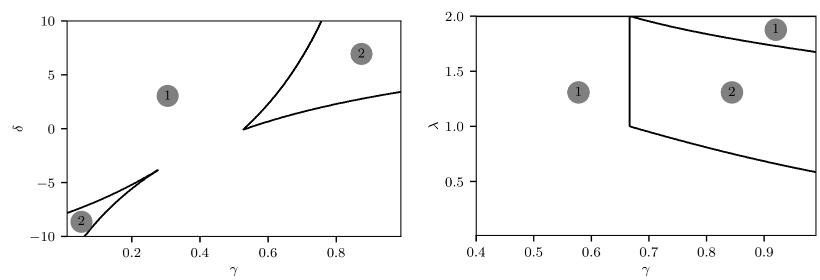
$\gamma = 0.7$, $\kappa_1 = \kappa_2 = 0.2$, $\lambda = 1$, $\mu_1 = 2$, $\mu_2 = 1$ and $\delta = 20$. There is a single mixed stable equilibrium. The same arrival rates and service parameters are used in figure 2(b), but we now prioritise the second queue: $\gamma = 0.2$. In this case, the only (stable) equilibrium is obtained for $\lambda_1 = 0.6$, for which both queues grow. Customers have an incentive to opt for the slower queue. Note that the faster queue grows as well as the service share of the slower queue grows with the slower queue size. In figure 2(c), we retained all parameters apart from the constant cost δ which equals $\delta = -10$ in this figure. We now have a single mixed equilibrium which is not stable. Just below and above the equilibrium solution, customers have an incentive to drift away from the equilibrium solution. Both $\lambda_1 = 0.6$ and $\lambda_1 = 1$ are equilibrium solutions, the former again corresponding to a non-ergodic queueing system as in figure 2(b). Finally, figure 2(d) shows a scenario with two mixed equilibria. The parameters are here chosen as follows: $\gamma = 0.22$, $\mu_1 = 1.82$, $\mu_2 = 2$, $\kappa_1 = \kappa_2 = 0.2$, $\lambda = 1.34$, and $\delta = -6$. From left to right, for $\lambda_1 = 0$ we have a stable equilibrium, the first mixed equilibrium is not stable, while the second is stable.

In Figure 3 we show how the number and the value of the equilibria vary with the priority γ. Figure 3a is plotted for the same parameter as figs 2(a) and (b) but with $\delta = 1$. There is only one stable equilibrium which increases as γ increases.



(a) (b) (c)

Fig. 3: Number and value of equilibria as a function of γ

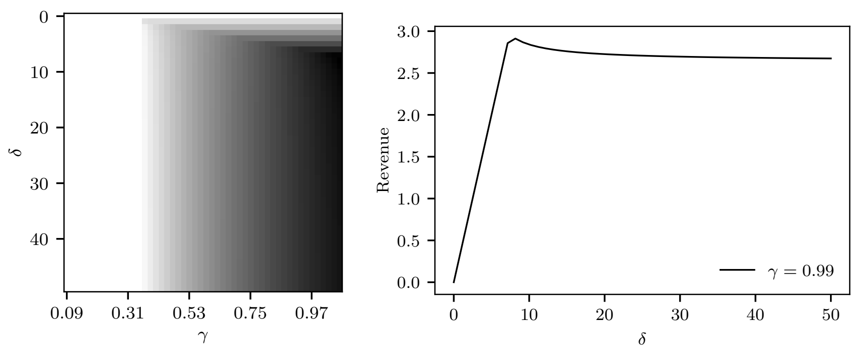


(a) (b)

Fig. 4: Number of stable and unstable equilibria for different (γ, δ) (a) and different (γ, λ) pairs.

Fig. 3b and 3c (resp.) are for parameters of figs. 2(c) and figs. 2(d)(resp.). In fig. 3b, there are initially three equilibria of which the mixed one is unstable. Further, $\lambda_1 = 1$ is an equilibrium for all values of γ. As priority increases the mixed equilibrium merges with the pure one at $\lambda_1 = 0.6$. On the other hand, in figure 3c there are two mixed equilibria one of which is stable and it increases towards $\lambda_1 = 1$ as γ increases. The mixed unstable equilibrium goes again merges with the pure one at $\lambda_1 = 0$.

The former figures show that the number of equilibria varies with the parameters. To further investigate parameter dependence, Figure 4 shows the regions with equal number of stable and unstable equilibria in the (γ, δ) (a) and (γ, λ) planes. We chose the following parameters in Figure 4(a): $\kappa_1 = \kappa_2 = 0.2$, $\mu_1 = 1.8$, $\mu_2 = 2$, $\lambda = 1.3$. In Figure 4(b), we set $\kappa_1 = \kappa_2 = 0$, $\mu_1 = 1$, $\mu_2 = 2$, and $\delta = 2$. For both figures, there is just one stable equilibrium and no unstable equilibria in region 1. In region 2 there are 2 stable equilibria as well as an unstable equilibrium.



(a) Revenue as a function of γ and δ (the darker the higher)

(b) Optimal choice of δ for $\gamma = 0.99$.

Fig. 5: Server-side optimization

Server-side optimisation: Finally, we investigate how the server should choose γ (the priority) and δ (payment by the customers) so as to maximise its revenue.

Let λ_m be the minimum value of λ_1 for which the system is stable. This value is given by (assuming $\mu_1 > \mu_2$): $\lambda_m = \max\left(0, \beta_2(\mu_1 - \mu_2)^{-1}\right)$. Note that at least a rate λ_m of customers have to go to queue 1 to keep the system stable. This means the server is guaranteed a revenue of $\lambda_m \delta$ and it can maximise revenue by setting an arbitrarily large δ since balking is not allowed in our model. In practice, faced with a high price, customers will either balk or go to another service provider. Since our model has only one server (a monopoly) and customers cannot balk, we impose a tax (or a penalty) of $\lambda_m \delta$ on the server. This tax (or penalty) is paid by the server to the society in order to keep its monopoly. Every customer choosing queue 1 pays δ but the server gets to keep only a fraction $(\lambda_1 - \lambda_m)/\lambda_1$ of this amount. The revenue of the server is then

$$(\lambda_1 - \lambda_m)\delta.$$

With this definition of the revenue, the server does not generate revenue when the customers that choose queue 1 are only those that are forced to make this choice in order to stabilise the system. The server can no longer take advantage of its monopoly and set arbitrarily high prices as seen in the example below.

We assume that the server is able to orient the Wardrop equilibrium chosen by the customers to the stable equilibrium of its choice. In figure 5a, the revenue is plotted as a heat map (the darker the higher) for different values of γ and δ. The other parameters are: $\kappa_1 = \kappa_2 = 0.2$, $\lambda = 1$, $\mu_1 = 2$, $\mu_2 = 1$. We observe that it is best for the server to give priority to class 1 (the faster class) by taking the largest γ. However, for $\gamma = 0.99$ (that is, full priority to class 1), there is an optimal value of δ that is in the interior as is shown in figure 5b.

4 Conclusion

We demonstrated that in a two-class discriminatory processor queue with an additional cost for joining the premium class, various equilibria can coexist depending on the system parameters. Notably, we found that certain conditions lead to non-ergodic behaviour, where the equilibrium traffic mix corresponds to a non-ergodic queueing system. This highlights the complexity of customer equilibria in the non-symmetric DPS queue. Additionally, we showed that the server can maximise revenue by adjusting the payment level of the higher priority class, thereby influencing the fraction of customers choosing the higher priority class.

References

1. I. Gurvich, M. A. Lariviere, and C. Ozkan. Coverage, coarseness, and classification: Determinants of social efficiency in priority queues. *Management Science*, 65(3):1061—1075, 2019.
2. Y. Hayel and B. Tuffin. Pricing for heterogeneous services at a discriminatory processor sharing queue. In *Proceedings of Networking*, 2005.
3. D. Fiems. Strategic revenue management for discriminatory processor sharing queues. In *Computer Performance Engineering and Stochastic Modelling (LNCS,volume 14231)*, pages 3–17. Springer, 2023.
4. G. Fayolle, I. Mitrani, and R. Iasnogorodski. Sharing a processor among many job classes. *Journal of the ACM*, 27(3):519–532, 1980.
5. L. Massoulié and J. Roberts. Bandwidth sharing: objectives and algorithms. *IEEE/ACM Transactions on Networking*, 10(3):320–328, 2002.
6. L. Massoulié and J. Roberts. Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems*, 15(1–2):185–201, 2000.
7. J. W. Roberts. A survey on statistical bandwidth sharing. *Comput. Networks*, 45(3):319–332, 2004.
8. S. F. Yashkov and A. S. Yashkova. Processor sharing: A survey of the mathematical theory. *Automation and Remote Control*, 68(9):1662–1731, 2007.
9. E. Altman, K. Avrachenkov, and U. Ayesta. A survey on discriminatory processor sharing. *Queueing Systems*, 53(1-2):53–63, 2006.
10. A. Orda, R. Rom, and N. Shimkin. Competitive routing in multiuser communication networks. *IEEE/ACM Transactions on Networking*, 1(5):510–521, 1993.
11. E. Altman and N. Shimkin. Individual equilibrium and learning in processor sharing systems. *Operations Research*, 46(6):776–784, 1998.
12. I. Ben-Shahar, A. Orda, and N Shimkin. Dynamic service sharing with heterogeneous preferences. *Queueing Systems*, 35(1–4):83–103, 2000.
13. L. Ravner, M. Haviv, and H. L. Vu. A strategic timing of arrivals to a linear slowdown processor sharing system. *European Journal of Operational Research*, 255:496–504, 2015.
14. E. Altman, U. Ayesta, and B. J. Prabhu. Load balancing in processor sharing systems. *Telecommunication Systems*, 47(1-2):35–48, 2011.
15. U. Ayesta, O. Brun, and B. J. Prabhu. Price of anarchy in non-cooperative load balancing games. *Performance Evaluation*, 68(12):1312–1332, 2011.
16. D. Fiems and B. J. Prabhu. Macroscopic modelling and analysis of flows during rush-hour congestion. *Performance Evaluation*, 149–150, 2021.
17. R. Hassin and M. Haviv. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*. Kluwer Academic, 2003.