Check for updates

Cooperative Game Theoretic Analysis of Shared Services

Anirban Mitra[1], Manu K. Gupta[1(✉)], and N. Hemachandra[2]

[1] Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India
{anirban_m,manu.gupta}@ms.iitr.ac.in
[2] Indian Institute of Technology Bombay, Mumbai 400076, Maharashtra, India
nh@iitb.ac.in

Abstract. Shared services are increasingly popular among firms and are often modeled as multi-class queuing systems. Several priority scheduling rules are possible to schedule customers from different classes. These scheduling rules can be static, where a class has strict priority over the other class, or can be dynamic based on delay and certain weights for each class. An interesting and important question is how to fairly allocate the waiting cost for shared services.

In this paper, we address the above problem using the solution concepts of cooperative game theory. We first appropriately define worth functions for each player (class), each coalition, and the grand coalition for multi-class M/G/1 queue with non-preemptive priority. It turns out that the worth function of the grand coalition follows Kleinrock's conservation law. We fully analyze the 2−class game and obtain the fair waiting cost allocations from several cooperative games' solution concepts viewpoints. These include Shapley value, the core, and nucleolus. We prove the 2−class game is convex which implies that the core is nonempty and the Shapley value allocation belongs to the core. Cooperative game-theoretic solutions capture fairness. We characterize the closed-form expression for these scheduling policies as bringing out various fairness aspects amongst scheduling policies. We consider Delay dependent priority (DDP) rule to determine fair scheduling policies from the Shapley value and the core-based allocation. We present extensive numerical experiments by partitioning the stability region for 2-class queues in three sub-regions.

Keywords: Cooperative game theory · Multi-class queueing systems · Dynamic priority scheduling · Shapley value · Nucleolus · The core · Achievable region · Delay dependent priority rule

1 Introduction

We consider the problem of fair scheduling in a queueing system where a server caters to several classes of customers. Customer's arrival to each class follows

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2023
Published by Springer Nature Switzerland AG 2023. All Rights Reserved
E. Hyytiä and V. Kavitha (Eds.): VALUETOOLS 2022, LNICST 482, pp. 33–47, 2023.
https://doi.org/10.1007/978-3-031-31234-2_3

Poisson process and a single server provide service to each class with non-preemptive priority and general service time distribution. In order to facilitate customer service, different priority scheduling rules are possible. These include static priority where one class of customers has strict high priority over the other class of customers. On the other hand, we can also have dynamic priority rules which may be dependent on the number in the queue or the delay of customers in each class. An important question in this setup is how the total cost can be fairly divided among its participants. This paper focuses on characterizing fair scheduling policies (fair waiting cost allocation) based on several solution concepts from cooperative game theory.

We first define an appropriate N-class game where the worth functions of each player, coalition, and the grand coalition are in terms of mean waiting time weighted by the load factors of each class. We then consider several solution concepts including the core, Shapely value, and nucleolus for fair waiting cost allocation. The core provides the set of coalitional and collectively rational allocations whereas the Shapley value is based on marginal utility and nucleolus minimizes the maximum dissatisfaction. We completely characterize these solution concepts for the 2-class game by first proving it to be convex and then deriving the closed-form expressions for allocations. We note that the class with a higher load factor results in higher Shapley value allocations. This is intuitive as a class that brings the higher load in the system gets allocated the higher waiting cost for a solution concept (Shapley value) which is based on marginal utility. For this game, Shapley value and nucleolus turn out to be the same. We also find the addition of Shapley value and the core allocations are the same as right hand side of Kleinrock's conservation law [1].

We consider Delay Dependent Priority (DDP) rule to determine fair scheduling policy. Delay dependent priority was first introduced by Kleinrock [2] and the mean waiting time in multi-class M/G/1 queue was derived under the DDP scheduling rule. A scheduling rule is *complete* if it covers all possible vectors of mean waiting time which are achievable. A set of all scheduling policies are obviously complete, however, one is interested in finding parameterized policies that are complete [3]. The completeness of DDP and its implications are discussed in [4]. DDP prioritizes each class based on the delay and a certain predefined weight parameter for each class. We now briefly explain the mechanism of DDP for two class queues. Let b_1 and b_2 be the weights associated with classes 1 and 2 respectively. Let $\beta := b_2/b_1$ be the ratio of these weights. Suppose $\beta = 0.75$, it means $b_2 = 0.75b_1$ i.e. the DDP priority scheduling prioritize queue class 2, 0.75 times than queue class 1. We utilize this delay dependent priority parameter and obtain fair scheduling policies with respect to the several solution concepts discussed above. We determine fair scheduling policy parameters associated with Shapley value ($\beta^{Shapley}$) and the core (β^{Core}).

We consider the entire region of stability for 2-class queues and decompose it appropriately in three sub-regions. In each sub-region, we identify the fair scheduling policy along with the closed-form expression ($\beta^{Shapley}$). This enables us to identify the sub-region where global FCFS should be the fair scheduling policy and also the sub-regions where a class should be given higher dynamic priority over the others. We present our understanding through numerical illustrations for different stability regions.

1.1 Related Literature

Several papers discussed the application of cooperative game theory to multi-class single server queues. These included either optimized service capacity [5–7] or fixed service capacity [8,9]. In an optimized service system of a single-server queueing game, each player is associated with a customer arrival stream and an M/M/1 queue operated by a group of players, which provides service to the members of arrival streams [10]. Gonzalez et al. [5] considered a cost-allocation problem of shared medical services. They assumed any coalition optimizes it's own service rate with respect to constraints of sojourn time. Under this assumption, they defined core allocation. Garcia-Sanz et al. [6] analyzed variations of the model proposed by Gonzalez et al. [5]. They considered more generic constraints of sojourn time and constraints on the mean waiting time of a queue. Yu et al. [7] considered game theoretic settings to analyze capacity sharing between different independent firms. Further, they modeled facilities as queueing systems. They showed conditions under which capacity sharing can be beneficial to the firms.

In the fixed service capacity of a single-server queueing game, each player has it's service capacity endowment and this can be modeled as a potential service rate [10]. Anily and Haviv [8] considered M/M/1 queueing systems and modeled it as a TU game when servers merge their capacities or each coalition of players pools these endowments into a single-server M/M/1 queueing setup. They analyzed the value of that coalition. Another paper Anily et al. [9] analyzed models with single-server situations for fixed network form. They also discussed redistribution of combined service capacity in network structure of M/M/1 queueing system. To the best of our knowledge, single server queues have not been analyzed for general service time distribution which we address in this paper. Another significant advancement in this paper includes the analysis of several game-theoretic solution concepts whereas literature often deals with a single solution concept [5,11,12].

Along with these several papers also discussed game-theoretic model in the context of resource pooling. Resource pooling has been analyzed by Armony et al. [13] and Liu et al. [11]. Armony et al. [13] developed game-theoretic model to evaluate the performance of pooling when servers strategically selects their capacity. Liu and Yu [11] also designed a multi-server queueing game and described shared services. They used the concept of Polymatroidal structure and found lower priority firms get subsidized by higher priority firms also sometimes lower priority firms can get net positive rewards which incentivizes them to collaborate. In these papers researchers mainly focused on shared services between independent firms. They modeled this scenario as a queueing system and analyzed them using cooperative game theory.

This paper is organized as follows: Sect. 2 describes the game theoretic model of N-class queueing system. Section 3 completely analyzes 2-class queueing games and several solutions. Section 4 provides a summary and discusses relevant future research directions.

2 Model Description

In this section, we discuss a cooperative game theoretic model for multi-class queues. Let's consider a multi-class M/G/1 queue with N number of classes. Some of the classes may have a higher priority over the other classes. We consider *non-preemptive* priority rules where customers with higher priority may not interrupt service of lower priority customers and higher priority class customers have to wait till the service completion of lower priority customers. And these priorities may be static (strict) or dynamic (based on the number or delay of each class). One of the possibilities for static priority is the case where class 1 has strict (non-preemptive) priority over class 2, class 2 has strict priority over class 3 and so on. There are $N!$ such permutations possible. Let Π denote the set of all such scheduling policies and $\mathbb{E}(W_i^\pi)$ be the mean waiting time of class i under the policy $\pi \in \Pi$.

Let the class i customers arrive according to the Poisson process with the rate $\lambda_i,\ i = 1, 2, \cdots, N$; let μ be the finite service rate for each class of customers. The load factor for class i will be $\rho_i = \lambda_i/\mu$ and $\rho = \sum_{i=1}^{N} \rho_i$. We model the above multi-class queue as N-player cooperative game, described below. In this game, each player corresponds to a class of multi-class queuing systems. We refer N player cooperative game as N-class game in this paper.

2.1 *N*-class Game

We model the N-class queueing system as a Transferable Utility (TU) game with each class being considered as a player. Let $(\mathcal{P}, v)$ be a TU game with $\mathcal{P} = \{1, \cdots, N\}$ as the set of players and $v : 2^N \to \mathbb{R}$ as the worth functions of players. $v(\{S\})$ represents the worth of coalition $S, S \subseteq \mathcal{P}$. From the properties of TU game, the worth function of the null set is zero, i.e., $v(\{\phi\}) = 0$. We define the worth function of player i (class i), the worth function of a possible coalition set $S \subseteq \mathcal{P}$, and the worth function of the grand coalition $\mathcal{G}$ as follows:

$$v(\{i\}) = \left(\rho_i \sum_{\pi \in \mathcal{N}_i} \mathbb{E}(W_i^\pi) \right), \tag{1}$$

$$v(\{S\}) = \frac{\sum_{i \in S} [\rho_i \sum_{\pi \in \mathcal{N}_S} \mathbb{E}(W_i^\pi)]}{|S|!}, S \subseteq \mathcal{P}, \tag{2}$$

$$v(\{\mathcal{G}\}) = \frac{\sum_{i \in N} [\rho_i \sum_{\pi \in \Pi} \mathbb{E}(W_i^\pi)]}{|\mathcal{G}|!}, \tag{3}$$

where $\mathcal{N}_i \subset \Pi$ are the set of policies such that player i has the highest priority and $\mathcal{N}_S$ will be the set of policies where players in set S have higher priority,

$\mathcal{N}_S = \cup_{\{i \in S\}} \mathcal{N}_i$. We now illustrate the scheduling policies in $\mathcal{N}_i, i = 1, \cdots, N$ and $\mathcal{N}_S$ for 2−class and 3−class games.

1. For a 2−class game $\mathcal{N}_1$ is a set of policies where player 1 has higher priority, $\mathcal{N}_1 = \{12\}$. Similarly, $\mathcal{N}_2$ will be $\{21\}$. Now consider the set $S = \{1, 2\}$, $\mathcal{N}_S$ is a set of policies where players 1 and 2 have higher priorities, $\mathcal{N}_S = \{12, 21\}$. Here policy 12 indicates that class 1 has higher static priority over class 2 and vice-versa for 21.
2. For 3−class game, $\mathcal{N}_1 = \{123, 132\}$, $\mathcal{N}_2 = \{213, 231\}$ and $\mathcal{N}_3 = \{312, 321\}$. Consider $S = \{1, 2\}$. Thus, $\mathcal{N}_S$ will be the set of policies such that class 1 or 2 has the highest priority. For this game, $\mathcal{N}_{\{1,2\}} = \{123, 132, 213, 231\}$. Similarly, $\mathcal{N}_{\{2,3\}} = \{213, 231, 312, 321\}$ and $\mathcal{N}_{\{3,1\}} = \{312, 321, 123, 132\}$.

2.2 Importance of Worth Functions in *N*-class Game

We determine the worth function of an individual class by the sum of mean waiting times under the policies where the class has the highest priority. This is because of the selfish nature of the individual players. Further, this sum is multiplied by the load factor of that class. The load factor multiplication ensures that a player can only value policies that give it an absolute priority.

Another important aspect of our worth function of a coalition ($S \subset \mathcal{G}$) is that we consider weighted average waiting time under the scheduling policies ($\pi \in \mathcal{N}_S$) where classes in a coalition are given higher priority. We are considering the highest priorities of players in the coalition because of the selfish nature of players in the coalition. The weights are given in terms of the load factor of the respective class. This is also in line with the worth function of the individual players.

For the grand coalition instead of taking participants of the coalition, we consider all the players in the game. Federgruen and Groenevelt [12] analyzed the performance of a queuing system by considering priority class queues where several leagues are formed. These leagues are different subsets of N. In the above game description, the leagues have similarities to the coalition. We now present the motivation behind the above worth functions. The grand coalition in Eq. (3) can be rewritten as follows since $|\Pi| = N!$.

$$v(\{\mathcal{G}\}) = \frac{\rho_1 \left[\mathbb{E}(W_1^{\pi_1}) + \cdots + \mathbb{E}(W_1^{\pi_{N!}})\right] + \cdots + \rho_N[(\mathbb{E}(W_N^{\pi_1}) + \cdots + \mathbb{E}(W_N^{\pi_{N!}})]}{|\mathcal{G}|!},$$

on rewriting the above expression by collecting the mean waiting time for each scheduling policy $\pi_1, \pi_2, \cdots, \pi_{N!}$, we have

$$v(\{\mathcal{G}\}) = \frac{\left[\rho_1 \mathbb{E}(W_1^{\pi_1}) + \cdots + \rho_N \mathbb{E}(W_N^{\pi_1})\right] + \cdots + \left[\rho_1 \mathbb{E}(W_1^{\pi_{N!}}) + \cdots + \rho_N \mathbb{E}(W_N^{\pi_{N!}})\right]}{|\mathcal{G}|!}.$$

From Kleinrock's conservation law [1], we have

$$\sum_{i=1}^{N} \rho_i \mathbb{E}(W_i^{\pi}) = \frac{\rho W_0}{1 - \rho} \ \forall \ \pi \in \Pi, \tag{4}$$

where, $W_0 = \sum_{i=1}^{N} \frac{\lambda_i}{2}[\sigma^2 + \frac{1}{\mu^2}]$ and σ^2 is variance of service time. We can now further simplify $v(\{\mathcal{G}\})$ using the above conservation law:

$$v(\{\mathcal{G}\}) = \frac{N!(\rho W_0)}{N!(1-\rho)} = \frac{\rho W_0}{1-\rho},$$

which is the right-hand side of Kleinrock's conservation law in Eq. (4). Therefore, we can say that the worth of the grand coalition is independent of scheduling policies.

We intend to explore several solution concepts from cooperative game theory and find out a fair scheduling policy for each solution concept in the above game. We now illustrate the results with the 2−class game in the next section.

3 2-class Game Illustration

In this section, we consider a 2−class game, i.e., $N = 2$. The set of players is $\mathcal{P} = \{1, 2\}$ and there are a total 2! scheduling policies. Let $\pi_1 = \{12\}$ and $\pi_2 = \{21\}$ be these two scheduling policies where player 1 (class 1) and player 2 (class 2) have strict priorities respectively. Note that $\Pi = \{\pi_1, \pi_2\}$. Let $v(\{1\})$, $v(\{2\})$ and $v(\{12\})$ are the worth functions for players 1, 2, and the grand coalition $\mathcal{G}$ respectively. By using the definition of N-class game from Eq. (1)-(3) worth functions of player 1, player 2 and, the grand coalition are as follows:

$$v(\{1\}) = \rho_1 \mathbb{E}(W_1^{12}), \tag{5}$$

$$v(\{2\}) = \rho_2 \mathbb{E}(W_2^{21}), \tag{6}$$

$$v(\{12\}) = \frac{\rho_1 \left[\mathbb{E}(W_1^{12}) + \mathbb{E}(W_1^{21})\right] + \rho_2 \left[\mathbb{E}(W_2^{12}) + \mathbb{E}(W_2^{21})\right]}{2}. \tag{7}$$

We now show below that the worth function of the grand coalition is higher than the addition of individual worth functions, i.e.,

$$v(\{12\}) > v(\{1\}) + v(\{2\}).$$

This implies that there is an incentive for players to collaborate. The mean waiting time expressions for class 1 and class 2 under policy $\{12\}$ and $\{21\}$ are as follows (see [14]):

$$\mathbb{E}(W_1^{12}) = \frac{W_0}{(1-\rho_1)} \quad \text{and} \quad \mathbb{E}(W_1^{21}) = \frac{W_0}{(1-\rho_2)(1-\rho)}, \tag{8}$$

$$\mathbb{E}(W_2^{21}) = \frac{W_0}{(1-\rho_2)} \quad \text{and} \quad \mathbb{E}(W_2^{12}) = \frac{W_0}{(1-\rho_1)(1-\rho)}. \tag{9}$$

Now, from the above equations we got the followings:

$$v(\{1\}) = \frac{\rho_1 W_0}{(1-\rho_1)},\ v(\{2\}) = \frac{\rho_2 W_0}{(1-\rho_2)},\ v(\{12\}) = \frac{\rho W_0}{(1-\rho)}.$$

We have,

$$v(\{1\}) + v(\{2\}) = \frac{\rho W_0 - \rho_1\rho_2 W_0}{(1-\rho) + \rho_1\rho_2}.$$

Therefore, $v(\{12\}) > v(\{1\}) + v(\{2\})$ as $\rho_1 > 0, \rho_2 > 0$ and $W_0 > 0$.

Convexity of 2-Class Game: We prove below that the 2−class game is convex. A TU game is convex if it follows supermodularity, which is

$$v(\{C \cup D\}) + v(\{C \cap D\}) \geq v(\{C\}) + v(\{D\}),\ \forall\ C,\ D \subseteq \mathcal{P}.$$

For the above 2−class game, we verify all the possible subsets of C and D ($C \subset D, D \subset C, C = D, C \cap D = \phi$). It follows that, $v(\{C \cup D\}) + v(\{C \cap D\}) \geq v(\{C\}) + v(\{D\})$. Therefore, we can say the 2−class game is convex. Now we explore the core in the next subsection.

3.1 The Core

The core is a solution concept from cooperative game theory which includes all the possible allocations which are coalitional and collectively rational. It follows from the convexity of the game that the core is non-empty [15] (see page 424). The following proposition presents the complete characterization of all allocations in the core for the 2−class game.

Proposition 1. *For 2-class game, the core, $\mathcal{C}(\mathcal{P}, v)$, is non-empty and a set given by:*

$$\mathcal{C}(\mathcal{P}, v) = \left\{(x_1, x_2) : \frac{\rho_1 W_0}{(1-\rho_1)} \leq x_1 \leq \frac{\rho_1 W_0}{(1-\rho_2)(1-\rho)},\ x_1 + x_2 = \frac{\rho W_0}{(1-\rho)}\right\}.$$

Proof. See Appendix A

Remark 1. *In the 2-class game, the addition of core allocations is right hand side of Kleinrock's conservation law, i.e., $x_1 + x_2 = \dfrac{\rho W_0}{1-\rho}$.*

We now explore the Shapley value solution concept for this 2−class game.

3.2 Shapley Value

Shapley value is an axiomatic solution concept in cooperative game theory that faithfully provides the marginal contribution of each player. The following proposition provides the Shapley value allocation for the 2−class game.

Proposition 2. *For the 2-class game, the Shapley value allocation for class 1, $\phi_1(v)$, and class 2, $\phi_2(v)$, are respectively given by*

$$\phi_1(v) = \frac{W_0\rho_1}{2}\left[\frac{\rho\rho_2 + 2(1-\rho)}{(1-\rho_1)(1-\rho_2)(1-\rho)}\right], \tag{10}$$

$$\phi_2(v) = \frac{W_0\rho_2}{2}\left[\frac{\rho\rho_1 + 2(1-\rho)}{(1-\rho_1)(1-\rho_2)(1-\rho)}\right]. \tag{11}$$

Proof. See Appendix A

Corollary 1. *In the 2-class game, the player with a higher load factor gets the higher Shapley value allocation and vice-versa.*

Proof. From Eqs. (10) and (11), we note that the Shapley values are dependent on the load factor of the respective and the other player. We also note that the denominator of Shapley values is the same and non-negative as $0 < \rho_1 < 1$ and $0 < \rho_2 < 1$ from the stability of queues. If $\rho_1 > \rho_2$ then $\phi_1 > \phi_2$. Similarly, if $\rho_2 > \rho_1$ then $\phi_2 > \phi_1$. Therefore, the corollary holds.

Corollary 2. *In the 2-class game, the addition of Shapley values is right hand side of the Kleinrock's conservation law, i.e.,* $\phi_1(v) + \phi_2(v) = \frac{\rho W_0}{1-\rho}$.

Proof. The result immediately follows from Proposition 2.

3.3 Fair Scheduling Policies

A fair scheduling policy is one that results in a fair mean waiting time for each class. The fairness of mean waiting times can be determined by the cooperative game-theoretic solution concepts. We exploit Remark 1 and Corollary 2 to obtain a fair scheduling policy from the core and Shapley value-based solution concepts respectively in the 2−class game.

Shapley Value-Based Fair Scheduling Policy. Note that the Kleinrock's conservation law for two classes is given by (under the scheduling policy π):

$$\rho_1 \mathbb{E}(W_1^{\pi}) + \rho_2 \mathbb{E}(W_2^{\pi}) = \frac{\rho W_0}{1-\rho}. \tag{12}$$

From Corollary 2, we have

$$\phi_1(v) + \phi_2(v) = \frac{\rho W_0}{1-\rho}. \tag{13}$$

The above expression can be rewritten as follows in view of the comparison of the above two.

$$\rho_1 \hat{\phi}_1(v) + \rho_2 \hat{\phi}_2(v) = \frac{\rho W_0}{1-\rho}, \tag{14}$$

where $\hat{\phi}_1(v) = \frac{W_0[\rho\rho_2 + 2(1-\rho)]}{2(1-\rho_1)(1-\rho_2)(1-\rho)}$ and $\hat{\phi}_2(v) = \frac{W_0[\rho\rho_1 + 2(1-\rho)]}{2(1-\rho_1)(1-\rho_2)(1-\rho)}$.

We are interested in exploring a scheduling policy $\pi^{Shapley}$ that schedules the customers in such a way that the mean waiting times for class 1 and class 2 under the scheduling policy $\pi^{Shapley}$ becomes $\hat{\phi}_1(v)$ and $\hat{\phi}_2(v)$ respectively. We exploit Delay dependent priority (DDP) queues to determine $\pi^{Shapley}$. DDP queues are the dynamic priority rules based on the delay of customers in queues [2]. The delay dependent priority scheduling is shown to be complete, i.e., it achieves all possible vectors of mean waiting times for 2−class queue (see [4]) under non-preemptive, non-anticipating and work conserving scheduling policy. Thus, we use DDP to find $\pi^{Shapley}$. The average waiting times for class 1 and class 2 under DDP scheduling policy is as follows [2,4,16–18]:

$$\mathbb{E}[W_1^{DDP}(\beta)] = \frac{W_0(1-\rho(1-\beta))}{(1-\rho)(1-\rho_1(1-\beta))}1_{\{\beta\leq 1\}} + \frac{W_0}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta}))}1_{\{\beta>1\}}, \tag{15}$$

$$\mathbb{E}[W_2^{DDP}(\beta)] = \frac{W_0}{(1-\rho)(1-\rho_1(1-\beta))}1_{\{\beta\leq 1\}} + \frac{W_0(1-\rho(1-\frac{1}{\beta}))}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta}))}1_{\{\beta>1\}}, \tag{16}$$

where $1_{\{.\}}$ is an indicator function and β is a parameter that determines the scheduling policy. $\beta := b_2/b_1$ where b_1 and b_2 are the weights associated with class 1 and class 2 respectively. $b_1 = 0$ implies strict higher priority to class 2 and it indicates $\beta = \infty$. On the other hand, $b_2 = 0$ implies strict higher priority to class 1 and it indicates $\beta = 0$. It has been shown in [4] that $0 \leq \beta \leq \infty$ achieves all possible vectors of mean waiting time, i.e., completeness of DDP. The following proposition finds the fair DDP scheduling policy parameter ($\beta^{Shapley}$) that schedules according to $\pi^{Shapley}$.

Proposition 3. *In the 2-class game, the DDP scheduling parameter* $\beta^{Shapley}$ *that achieves the fair scheduling policy* $\pi^{Shapley}$ *is given by:*

$$\beta^{Shapley} = \frac{(2-\rho)(1-\rho_1)}{\rho\rho_1 + 2(1-\rho)}1_{\{\rho_1\geq\rho_2\}} + \frac{\rho\rho_2 + 2(1-\rho)}{(2-\rho)(1-\rho_2)}1_{\{\rho_1<\rho_2\}}. \tag{17}$$

Proof. See Appendix A

We have the following interpretation of the above proposition.

1. When load factors are equal, i.e., $\rho_1 = \rho_2$, $\beta^{Shapley}$ turns out to be 1. This implies global FCFS scheduling policy.
2. When load factor for class 1 is higher than that of class 2, i.e., $\rho_1 > \rho_2$, $0 < \beta^{Shapley} < 1$. This implies class 1 has higher dynamic priority than class 2.
3. When the load factor for class 2 is higher than that of class 1, i.e., $\rho_2 > \rho_1$, $1 < \beta^{Shapley} < \infty$. This implies class 2 has a higher dynamic priority than class 1.
4. At the extreme case when class 1 achieves the highest load factor ($\rho_1 \to 1$) then $\beta^{Shapley} \to 0$. This implies static high priority to class 1.
5. At another extreme case when class 2 achieves the highest load factor ($\rho_2 \to 1$) then $\beta^{Shapley} \to \infty$. This implies static high priority to class 2.

In Fig. 1, we discuss the fair scheduling policy allocation based on Shapley value for the entire stability region ($\rho_1 + \rho_2 < 1$). This stability region is divided into three sub-regions: 1) $\rho_1 > \rho_2$ 2) $\rho_1 < \rho_2$ 3) $\rho_1 = \rho_2$. The stability region

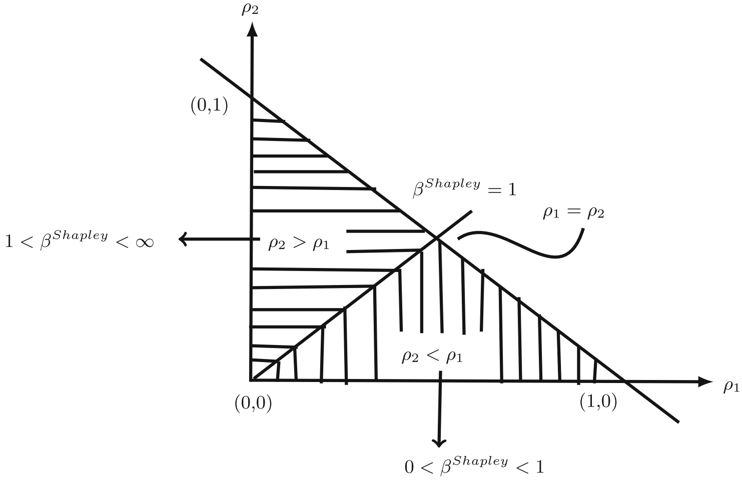

Fig. 1. Fair scheduling policy based on Shapley value, $\beta^{Shapley}$, within the stability region ($\rho_1 + \rho_2 < 1$)

$\rho_1 > \rho_2$ is shown by the bottom triangle. In this sub-region, $0 < \beta^{Shapley} < 1$, indicates higher dynamic priority to class 1. Another sub-region ($\rho_1 < \rho_2$) indicates the upper triangle in Fig. 1. Inside this sub-region ($1 < \beta^{Shapley} < \infty$), indicates higher dynamic priority to class 2. The third sub-region indicates a line ($\rho_1 = \rho_2$). In this sub-region, $\beta^{Shapley} = 1$ which implies global FCFS scheduling policy.

We now present the numerical experiments to illustrate the above. In Fig. 2, we illustrate the above point (1) where both the classes have equal load factors ($\rho_1 = \rho_2$). In this case, we observe $\beta^{Shapley} = 1$ and it coincides with global FCFS scheduling policy. The parameter settings are mentioned in the caption. In all the figures, square and cross represent the scheduling policies corresponding to $\beta^{Shapley}$ and global FCFS respectively. Figure 3 illustrates the above point (2) where $\rho_1 > \rho_2$. In this case, $\beta^{Shapley}$ turns out to be 0.54 which is between 0 and 1 as expected. Note that the mean waiting time corresponding to the global FCFS scheduling policy is shifting towards policy 21 ($\beta = \infty$). This is expected as the implication of increased load factor from class 1 implies global FCFS to behave in the same way as policy 21. Figure 4 illustrates the case where $\rho_2 > \rho_1$. In this case, $\beta^{Shapley}$ turns out to be 3.24 which is larger than 1 as expected. Note that the mean waiting time corresponds to the scheduling policy $\beta^{Shapley}$ is the middle point of the waiting times corresponding to policies 12 and 21. This can be verified mathematically from the expression of $\hat{\phi}_i(v), i = 1, 2$, and the mean waiting time expressions in (8) and (9).

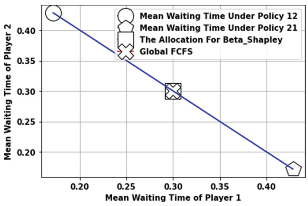

Fig. 2. Policy illustration for $\rho_1 = \rho_2$. Parameter settings: $\lambda_1 = \lambda_2 = 2$ and $\mu = 5$.

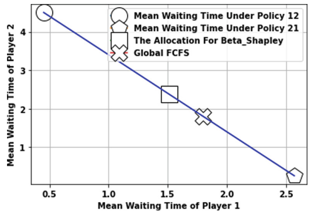

Fig. 3. Policy illustration for $\rho_1 > \rho_2$ Parameter settings: $\lambda_1 = 3,\ \lambda_2 = 1$ and $\mu = 5$.

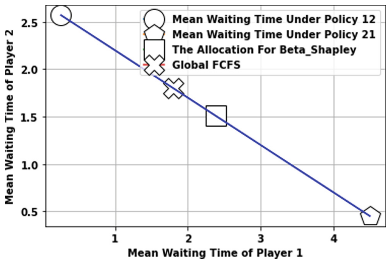

Fig. 4. Policy illustration for $\rho_1 < \rho_2$. Parameter settings: $\lambda_1 = 1, \lambda_2 = 3.8$ and $\mu = 5$.

Core Based Fair Scheduling Policy. Note that Kleinrock's conservation law for two classes is given by (under the scheduling policy π):

$$\rho_1 \mathbb{E}(W_1^\pi) + \rho_2 \mathbb{E}(W_2^\pi) = \frac{\rho W_0}{1-\rho}. \tag{18}$$

From Remark 1, we have

$$x_1 + x_2 = \frac{\rho W_0}{1-\rho}. \tag{19}$$

The above expression can be rewritten as follows in view of the comparison of the above two.

$$\rho_1 \hat{x}_1 + \rho_2 \hat{x}_2 = \frac{\rho W_0}{1-\rho}, \tag{20}$$

where $\frac{W_0}{(1-\rho_1)} \leq \hat{x}_1 \leq \frac{W_0}{(1-\rho_2)(1-\rho)}$ and $\hat{x}_2 = \frac{1}{\rho_2} \frac{\rho W_0}{(1-\rho)} - \frac{\rho_1 \hat{x}_1}{\rho_2}$. We are interested in exploring a scheduling policy π^{Core} that schedules the customers in such a way that the mean waiting times for class 1 and class 2 under the scheduling policy π^{Core} becomes $\hat{x}_1$ and $\hat{x}_2$ respectively. We again consider DDP scheduling policy since it is complete. The following proposition finds the range of fair DDP scheduling policy parameter (β^{Core}) that schedules according to π^{Core}.

Proposition 4. *In the 2-class game, the range of DDP scheduling parameter $\beta^{Core} \in [0, \infty]$ that achieves the fair scheduling policy π^{Core}.*

Proof. See Appendix A

3.4 Nucleolus

Nucleolus is a solution concept in cooperative game theory which minimizes the maximum excess (unhappiness) of the most unhappy coalition, the second most unhappy coalition, and so on. Since the core is non-empty (see proposition 1), nucleolus belongs to the core and it is unique [15] (see page 455).

Proposition 5. *For the 2-class game nucleolus is the Shapley value.*

Proof. See Appendix A

4 Discussion

We considered a generic multi-class M/G/1 queuing system and designed a cooperative game theoretic setup to obtain a fair scheduling policy. We completely characterized the 2-class game using solution concepts such as Shapley value, the core, and nucleolus. We showed Shapley value and nucleolus to be the same for 2−class game. Also, the waiting cost allocations by the Shapley value and the core respect Kleinrock's conservation law.

The major contribution of our paper is to determine fair scheduling policy by exploiting Kleinrock's delay dependent priority for 2−class queues. We partition the entire stability region into three parts: 1) $\rho_1 > \rho_2$, 2) $\rho_1 < \rho_2$ and 3) $\rho_1 = \rho_2$. And obtain a fair scheduling scheme in each sub-region. We obtain a closed-form expression of fair scheduling policy parameter associated with Shapley value ($\beta^{Shapley}$) and a range of β associated with the core (β^{Core}).

This model can be extended in several ways. Firstly, we can extend this model to N-class queueing system with different service rates/distributions for each class and explore a fair scheduling policy for this. Secondly, we can consider a queueing network, where multiple servers are present with multiple queue classes. In this complex scenario, obtaining a fair scheduling policy is another interesting future avenue.

A Proofs

Proof of Proposition 1: It follows that for any convex game the core is non-empty and the nucleolus belongs to the core [15] (See page 424 and 455). As we already proved this is a Convex Game. so, the core is non-empty and the nucleolus belongs to the core. Also, Shapley value belongs to the core. From [15] (see page 416) the core, $\mathcal{C}(\mathcal{P}, v) = \{x = (x_1, \cdots, x_n) \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = v(\{N\}); \sum_{i \in C} x_i \geq$

$v(\{C\}), \forall\ C \subseteq \mathcal{P}\}$. For 2−class game the core is $\mathcal{C}(\mathcal{P}, v) = \{x = (x_1, x_2) \in \mathbb{R}^2 : \sum_{i=1}^{2} x_i = v(\{N\}); \sum_{i \in C} x_i \geq v(\{C\}), \forall\ C \subseteq \mathcal{P}\}$. For this 2−class game by using Eqs. (1)-(3) we simplify write $x_1 \geq \frac{\rho_1 W_0}{(1-\rho_1)}, x_2 \geq \frac{\rho_2 W_0}{(1-\rho_2)}, x_1 + x_2 \geq \frac{\rho W_0}{(1-\rho)}, x_2 = \frac{\rho W_0}{(1-\rho)} - x_1$. From these expressions we get Proposition 1. ■

Proof of Proposition 2: It follows that there exists exactly one mapping $\phi : \mathbb{R}^{2^N-1} \rightarrow \mathbb{R}^N$ that satisfies all three Axioms (Symmetry, Linearity, and Carrier) [15] (see page 432). This mapping satisfies

$$\phi_i(v) = \sum_{C \subseteq \mathcal{P} \setminus \{i\}} \frac{|C|!(N-|C|-1)!}{N!} \Big\{ v(C \cup \{i\}) - v(C) \Big\}\ \forall\ i \in \mathcal{P},\ v \in \mathbb{R}^{2^N-1}$$

where N is the total number of players. The above expression for $\phi_i(v)$ gives the expected contribution of player i to the worth of any coalition and is called Shapley value. For this 2−class game all the possible coalitions are $(\{\phi\}), (\{1\}), (\{2\})$ and $(\{12\})$. From the above expression, the Shapley values of player 1 and player 2 are:

$$\phi_1(v) = \frac{1}{2}\left[v(\{1\}) + v(\{12\}) - v(\{2\})\right],$$

$$\phi_2(v) = \frac{1}{2}\left[(v(\{2\}) + v(\{12\}) - v(\{1\})\right].$$

Now from Eqs. (1, 2, 3) and mean waiting time expressions we simplify the Shapley value of player 1 and player 2 as given in the proposition. ■

Proof of Proposition 3: Using Eqs. (18) and (20) we can write $\hat{\phi}_1 = \mathbb{E}(W_1^\beta)$ and $\hat{\phi}_2 = \mathbb{E}(W_2^\beta)$. From Eqs. (15) and (16) mean waiting time for class 1 and class 2 under DDP are as follows:

$$\mathbb{E}[W_1^{DDP}(\beta)] = \frac{W_0(1-\rho(1-\beta))}{(1-\rho)(1-\rho_1(1-\beta))} 1_{\{\beta \leq 1\}} + \frac{W_0}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta}))} 1_{\{\beta > 1\}},$$

$$\mathbb{E}[W_2^{DDP}(\beta)] = \frac{W_0}{(1-\rho)(1-\rho_1(1-\beta))} 1_{\{\beta \leq 1\}} + \frac{W_0(1-\rho(1-\frac{1}{\beta}))}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta}))} 1_{\{\beta > 1\}},$$

Now by comparing $\hat{\phi}_1$ and $\hat{\phi}_2$ with the mean waiting time of class 1 and class 2 under DDP for $\beta \leq 1$ we get the following:

$$\frac{W_0[\rho\rho_2 + 2(1-\rho)]}{2(1-\rho_1)(1-\rho_2)(1-\rho)} = \frac{W_0(1-\rho(1-\beta^{Shapley}))}{(1-\rho)(1-\rho_1(1-\beta^{Shapley}))}$$

and

$$\frac{W_0[\rho\rho_1 + 2(1-\rho)]}{2(1-\rho_1)(1-\rho_2)(1-\rho)} = \frac{W_0}{(1-\rho)(1-\rho_1(1-\beta^{Shapley}))}$$

Similarly, for $\beta > 1$ we get the following:

$$\frac{W_0[\rho\rho_2 + 2(1-\rho)]}{2(1-\rho_1)(1-\rho_2)(1-\rho)} = \frac{W_0}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta^{Shapley}}))}$$

and

$$\frac{W_0[\rho\rho_1 + 2(1-\rho)]}{2(1-\rho_1)(1-\rho_2)(1-\rho)} = \frac{W_0(1-\rho(1-\frac{1}{\beta^{Shapley}}))}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta^{Shapley}}))}$$

Compiling all the results we get the outcome mentioned in Proposition 3. ■

Proof of Proposition 4: From Eq. (15) and (16) mean waiting time for class 1 and class 2 under DDP are as follows:

$$\mathbb{E}[W_1^{DDP}(\beta)] = \frac{W_0(1-\rho(1-\beta))}{(1-\rho)(1-\rho_1(1-\beta))}1_{\{\beta\le 1\}} + \frac{W_0}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta}))}1_{\{\beta>1\}},$$

$$\mathbb{E}[W_2^{DDP}(\beta)] = \frac{W_0}{(1-\rho)(1-\rho_1(1-\beta))}1_{\{\beta\le 1\}} + \frac{W_0(1-\rho(1-\frac{1}{\beta}))}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta}))}1_{\{\beta>1\}},$$

For the 2-class game The Core expression we got, $x_1 \ge \dfrac{\rho_1 W_0}{(1-\rho_1)}, x_2 \ge \dfrac{\rho_2 W_0}{(1-\rho_2)}, x_1 + x_2 \ge \dfrac{\rho W_0}{(1-\rho)}, x_2 = \dfrac{\rho W_0}{(1-\rho)} - x_1$. Now by comparing the mean waiting time for class 1 under DDP with Core allocation of class 1 (x_1) and mean waiting time for class 2 under DDP with Core allocation of class 2 (x_2) we get the following:

$$\frac{W_0}{(1-\rho)(1-\rho_1(1-\beta^{Core}))} \ge \frac{W_0}{(1-\rho_1)}, \beta^{Core} \le 1$$

$$\frac{W_0}{(1-\rho)(1-\rho_2(1-\frac{1}{\beta^{Core}}))} \ge \frac{W_0}{(1-\rho_2)}, \beta^{Core} > 1$$

After simplifying the above inequalities we get the ranges of β as mentioned in this proposition. ■

Proof of Proposition 5: Let's assume Nucleolus is Shapley value (ϕ_1, ϕ_2). Now excess of coalition 1, $e(1, \phi_1) = v(1) - \phi_1$ and excess of coalition 2, $e(2, x) = v(2) - \phi_2$. Excess of the grand coalition, $e(12, \phi) = 0$. We found $e(1, \phi_1) = e(2, \phi_2) = \dfrac{-\rho_1\rho_2 W_0(2-\rho)}{2(1-\rho_1)(1-\rho_2)(1-\rho)}$. Now suppose there are any other allocations (x_1^*, x_2^*), that can further reduce $e(1, \phi_1)$ and $e(2, \phi_2)$. It means $v(\{1\}) - x_1^* < v(\{1\}) - \phi_1$ and $v(\{2\}) - x_2^* < v(\{2\}) - \phi_2$. This implies $x_1^* > \phi_1$ and $x_2^* > \phi_2$ which cannot be true as $\phi_1 + \phi_2 = \dfrac{\rho W_0}{1-\rho}$. Thus Shapley value is Nucleolus.

References

1. Kleinrock, L.: A conservation law for a wide class of queueing disciplines. Naval Res. Logistics Q. **12**(2), 181–192 (1965)
2. Kleinrock, L.: A delay dependent queue discipline. Naval Res. Logistics Q. **11**(3–4), 329–341 (1964)
3. Mitrani, I., Hine, J.H.: Complete parameterized families of job scheduling strategies. Acta Informatica **8**(1), 61–73 (1977)
4. Gupta, M.K., Hemachandra, N., Venkateswaran, J.: Some parameterized dynamic priority policies for two-class M/G/1 queues: completeness and applications. ACM Trans. Model. Perform. Eval. Comput. Syst. (TOMPECS) **5**(2), 1–37 (2020)
5. González, P., Herrero, C.: Optimal sharing of surgical costs in the presence of queues. Math. Methods Oper. Res. **59**(3), 435–446 (2004)
6. García-Sanz, M.D., Fernández, F.R., Fiestras-Janeiro, M.G., García-Jurado, I., Puerto, J.: Cooperation in markovian queueing models. Eur. J. Oper. Res. **188**(2), 485–495 (2008)
7. Yu, Y., Benjaafar, S., Gerchak, Y.: Capacity sharing and cost allocation among independent firms with congestion. Prod. Oper. Manag. **24**(8), 1285–1310 (2015)
8. Anily, S., Haviv, M.: Cooperation in service systems. Oper. Res. **58**(3), 660–673 (2010)
9. Anily, S., Haviv, M.: Homogeneous of Degree One Games are Balanced with Applications to Service Systems. Tel Aviv University, Faculty of Management, The Leon Recanati Graduate (2011)
10. Karsten, F., Slikker, M., Van Houtum, G.-J.: Resource pooling and cost allocation among independent service providers. Oper. Res. **63**(2), 476–488 (2015)
11. Liu, H., Yu, Y.: Incentives for shared services: multiserver queueing systems with priorities. Manuf. Serv. Oper. Manag. **24**(3), 1751–1759 (2022)
12. Federgruen, A., Groenevelt, H.: M/G/c queueing systems with multiple customer classes: characterization and control of achievable performance under nonpreemptive priority rules. Manage. Sci. **34**(9), 1121–1138 (1988)
13. Armony, M., Roels, G., Song, H.: Pooling queues with strategic servers: the effects of customer ownership. Oper. Res. **69**(1), 13–29 (2021)
14. Gelenbe, E., Mitrani, I.: Analysis and Synthesis of Computer Systems, vol. 4. World Scientific, Singapore (2010)
15. Narahari, Y.: Game Theory and Mechanism Design, vol. 4. World Scientific, Singapore (2014)
16. Kleinrock, L.: Queueing Systems, vol. 1. Wiley-Interscience, Hoboken (1975)
17. Kanet, J.J.: A mixed delay dependent queue discipline. Oper. Res. **30**(1), 93–96 (1982)
18. Sinha, S.K., Rangaraj, N., Hemachandra, N.: Pricing surplus server capacity for mean waiting time sensitive customers. Eur. J. Oper. Res. **205**(1), 159–171 (2010)