Operations Research
Vol. 58, No. 3, May–June 2010, pp. 660–673
ISSN 0030-364X | EISSN 1526-5463 | 10 | 5803 | 0660
DOI 10.1287/opre.1090.0737

Cooperation in Service Systems

Shoshana Anily
Faculty of Management, Tel Aviv University, Tel Aviv 69978, Israel, anily@post.tau.ac.il

Moshe Haviv
Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel,
haviv@mscc.huji.ac.il

We consider a number of servers that may improve the efficiency of the system by pooling their service capacities to serve the union of the individual streams of customers. This economies-of-scope phenomenon is due to the reduction in the steady-state mean total number of customers in the system. The question we pose is how the servers should split among themselves the cost of the pooled system. When the individual incoming streams of customers form Poisson processes and individual service times are exponential, we define a transferable utility cooperative game in which the cost of a coalition is the mean number of customers (or jobs) in the pooled system. We show that, despite the characteristic function is neither monotone nor concave, the game and its subgames possess nonempty cores. In other words, for any subset of servers there exist cost-sharing allocations under which no partial subset can take advantage by breaking away and forming a separate coalition. We give an explicit expression for all (infinitely many) nonnegative core cost allocations of this game. Finally, we show that, except for the case where all individual servers have the same cost, there exist infinitely many core allocations with negative entries, and we show how to construct a convex subset of the core where at least one server is being paid to join the grand coalition.

Subject classifications: queues; games/group decisions: bargaining; cooperative.
Area of review: Manufacturing, Service, and Supply Chain Operations.
History: Received July 2007; revisions received December 2007, June 2008, December 2008, May 2009; accepted May 2009. Published online in *Articles in Advance* December 29, 2009.

1. Introduction

Many retailing, manufacturing, and service systems consist of entities (firms or service providers) that may benefit from collaboration if economies of scope prevail. This phenomenon is well known in supply chain management, where, for example, retailers coordinate their replenishment from a supplier to save on delivery costs, and where manufacturers coordinate the operations of various stages in the production process to save on the holding costs. In general, coordination enables a better exploitation of the system's resources, which in turn reduces the total cost. Once the operational policy of the supply chain or service system is determined, the next natural question to be asked is how to allocate the total cost among the various entities that compose the system. In this research we pose similar questions in service management systems in which collaboration among service providers is beneficial to the system as a whole and to employ principles from cooperative game theory to answer them.

There are various applications where economies of scope prevail in service systems. For example, imagine a manufacturer of large electronic appliances such as refrigerators, microwaves, washing machines, and dryers who offers service contracts for repairs to customers that purchase a new product. Each of these functions has its demand rate. Suppose also that the manufacturer trains a number of technicians in each selling zone. In one extreme case, each technician is trained to repair just one type of product, whereas a better service, with respect to waiting times and length of waiting lines, can be obtained by training the technicians to repair all types of products. It makes sense to pool all servers so that all of them will serve all types of requests. In case that training is needed (as in the manufacturer management example above), this will come with some fixed costs. However, these costs are transient and much will be saved in the long run.

The obvious gain is that there will be no queue for one server while another is idle. This gain is translated into a reduction in the mean queueing time, or equivalently, by Little's law, to a reduction in the mean (total) queue length.

The issue of *cost allocation* or *cost sharing* among participants is in the heart of *transferable utility cooperative game theory*. In this model there is a cost associated with each subset, or *coalition*, of participants, usually the cost inflicted on them when they act alone. Moreover, in many cases it is socially optimal for all of them to fully cooperate, i.e., to form the *grand coalition*. In such cases the main question is how the participants should share this social cost among themselves in a fair or/and justified manner, taking into consideration, for example, both fairness and the relative power of each of them.

The problem we consider here is that of service providers, each of which needs to serve its own stream of customers. Thus, each service provider is modeled as an $M/M/1$ service system that is characterized by its own Poisson arrival rate and its own potential exponential service rate. Cooperation among subsets of service providers is possible, in which case we assume that the joint entity becomes a single server $M/M/1$ system with an arrival rate that is the sum of the respective individual rates, and a potential service rate value that is a concave-increasing function of the sum of the individual service rates.

The cost of an individual customer in a queue is usually measured by his/her mean waiting time. However, when one looks into the cost of a class of users, this mean waiting time needs to be multiplied by the arrival rate of that class (to get the corresponding cost per unit of time). By Little's law, this product is the mean number in the system of this class of customers. Hence, when we define below the cost of a coalition of servers (or firms), we look at the resulting mean queue length once this coalition is formed.

Cooperation in a service system that consists of a number of individual $M/M/1$ service providers makes sense because congestion in the system is smaller in the case of cooperation than the corresponding sum when they act separately. We will show that the social gains, measured by the reduction of the congestion in comparison with the prepooling case, are maximized when the grand coalition is formed.

As said above, the next question is: How do the service providers share among themselves the total cost of congestion when the grand coalition is formed? It is clear that some contribute more to the partnership, whereas others contribute less. For example, a service provider with a small arrival rate but large service capacity contributes less to the congestion than a service provider with a large arrival rate but small service capacity. Hence, the former should bear only a small fraction of the total costs in comparison with what the latter needs to be charged with. In fact, even the option that the latter will pay the former (making the net contribution of the former towards the joint cost negative) to persuade him, the former, to cooperate should not be ruled out a priori.

Our approach to this question is closely related to the one in Anily and Haviv (2007). There, we looked into an inventory model in which fixed setup costs for placing orders lead to cooperation among retailers, which results in placing joint orders. The steps we take are as follows. We first define a transferable utility cooperative game whose set of players is the service providers and where the cost associated with each coalition, known as the characteristic function, is the mean number of customers in an $M/M/1$ system formed by a single server with the union of the respective arrival streams and a service rate that is an increasing concave function of the respective sum of potential service rates and of arrival rates. Because the proposed game turns out to be *subadditive*, the cost of the grand coalition does not exceed the sum of the costs over any partition of the service providers. This phenomenon makes the grand coalition the optimal social formation and hence a natural end to a reasonable bargaining process. Then, we look for the core of the game, namely, for cost allocations such that no subset of players has a way to form a subcoalition whose cost is smaller than the sum of what is allocated to its members. A closer analysis of the characteristic function revealed that the suggested game, in contrast to the inventory game considered in Anily and Haviv (2007), is not concave. Had the game been concave (as in Anily and Haviv 2007), then the nonemptiness of the core would have been guaranteed. By using a different approach, we show that the core of the game under consideration is not empty (on top of being a convex set, which is always the case). In addition, we identify the set of all nonnegative core cost allocations of the game. In particular, we are able to state all extreme points of this convex set explicitly. Also, we show that the core is a singleton only in the trivial case where only one server exists. Otherwise, there are infinitely many nonnegative core allocations. The above results are achieved by defining an auxiliary concave game whose core is identical to the subset of the core of the original game containing only nonnegative vectors, namely, allocations where no server gets paid for cooperation. The characterization of that subset of the core is done by using the fact that there exist closed-form expressions for the extreme points of the core of a concave game. Finally, we show that with the exception of the trivial case where all utilization levels across individual servers coincide, there exist infinitely many core cost allocations in which at least one entry is negative. Such negative entries correspond to servers who are being paid by the other servers to make them join the grand coalition. In particular, we explicitly identify the set of servers that are never paid for joining the grand coalition, i.e., these are the congested servers that are allocated nonnegative costs in all core allocations. For each of the other servers we provide an infinite number of core allocations in which that server is paid.

There exists a literature on the pooling of servers, but to the best of our knowledge there is no extensive literature on the application of cooperative game theory to queueing systems leading to cost allocation mechanisms. One exception is Yu et al. (2006), where a model in which service capacity comes with a linear cost of its rate is considered. The service providers need to share the cost of the optimal pooled rate among themselves. In Yu et al. (2006) an imputation, i.e., a cost-sharing mechanism where everybody is better off than acting individually, is given. Another exception is Granot and Sosic (2003). There, service providers, each of which has its own quality of service (measured by guaranteed mean waiting time), cooperate by having a single server and applying a priority scheme so as to meet the individual requirements. However, they need to share the joint cost (linear in service capacity). They show that an allocation proportional to the individual's arrival rate is a core allocation. Finally, we would like to mention a related paper by Gonzalez and Herrero (2004), where the focus is on the Shapley value.

There are a few papers we are aware of where customers are treated as the players, and the question is how to share the total waiting costs among themselves. Katta and Sethuraman (2006), Chun (2006a, b), and Maniquet (2003) deal with finite population models (making them more of scheduling problems rather than queueing problems), whereas Haviv (1999) and Haviv and Ritov (1998) deal with an infinite stream of arrivals to an M/G/1 queue and charging each one of the customers, or groups of them, in a way that shares the total cost among them. The paper by Haviv and Ritov (1998) is a more adhoc and tailor-made approach to queueing models, whereas Haviv (1999) applies the Aumann-Shapley cost allocation mechanism, which suits continuous models in which each player (in this case, customer) has an infinitesimal value.

Pooling of resources in supply chains has attracted much more attention from researchers. Splitting of gains due to consolidation of orders in the multistore economic order quantity (EOQ) with safety stock is analyzed in Gerchak and Gupta (1991). Later, in Hartman and Dror (1996) and Ozen et al. (2004), this problem is cast in a cooperative game framework, and various solution concepts as the core and the Shapley value are considered. Hartman et al. (2000) and Muller et al. (2002) were the first to consider the issue of coordinating orders and sharing the cooperation gains in the multiretailer newsvendor problem. In Slikker et al. (2005), a more general variant of the model introduced in Hartman et al. (2000) and Muller et al. (2002) is considered, in which the retailers may hold separate inventories; thus, transshipment costs may be involved. The papers show that the core of any general newsvendor game of this type is nonempty. In Ozen et al. (2004), the benefit of pooling the inventories at the retailers after demand realization is analyzed in a more complex supply chain. The core of the associated cooperative game is shown to be nonempty. Recently, some papers have assumed that ordering decisions of the retailers are made independently/competitively, whereas the transshipment/allocation decisions, which are made after observing the demands, are taken cooperatively; see Anupindi et al. (2001), Granot and Sosic (2003), and Rudi et al. (2001).

In the context of the joint replenishment problem (JRP), Meca et al. (2004) considers the cost allocation problem in the special case where all the minor setup costs are zero. Later, in Anily and Haviv (2007), general cost parameters are allowed. In Dror and Hartman (2005) a JRP similar to the one in Anily and Haviv (2007) is considered, but with a different definition of the characteristic function, which results, in particular, in a nonconcave game. In Zhang (2009) a generalization of Anily and Haviv (2007) is proposed for the JRP of a one-warehouse multiretailer model with central inventory, and an ordering cost function that is a general submodular set function. The author defines the problem as a cooperative game, and shows that its core is nonempty by invoking a strong duality theorem on a mathematical nonconvex formulation of the optimization problem.

Another stream of research in supply chains deals with the multiretailer economic lot-sizing (ELS) model. Here, depending on the cost structure, if economies of scale prevail, it may be beneficial to the retailers to cooperate and place joint orders. The analysis of variants of this problem are considered in Guardiola et al. (2009) and van den Heuvel et al. (2007). An analysis of a variant that introduces periodic general concave ordering cost functions and backlogging is provided in Chen and Zhang (2009). As in van den Heuvel et al. (2007), it is assumed that the retailers share the same cost parameters. By invoking LP duality on an integer programming formulation of the problem, it is shown that its LP relaxation has a zero duality gap. Similarly to the results in Zhang (2009), the optimal dual solution is shown to be in the core.

The rest of this paper is organized as follows. Section 2, introduces the notations, states the model formally, claims the nonemptiness of the core, and presents some basic results, including the fact that the resulting game is not concave. In §3, we analyze the nonnegative core allocations by defining an auxiliary concave game whose extreme points are stated explicitly. We further show that the core of the auxiliary game coincides with the nonnegative subset of the core of the original game, meaning that we provide an explicit expression for all nonnegative core allocations for the original game. Then, we show that with the exception of the trivial case where there is only one server, there exist an infinite number of nonnegative core allocations. The proof of the concavity of the auxiliary game is deferred to the appendix. In §4, we show that with the exception of another trivial case where all servers have identical utilization levels, there exist infinitely many core allocations in which at least one of the servers is being paid by the other servers. We also design a constructive method that generates such an infinite subset of the core. Finally, §5 concludes with some extensions and suggestions for future research.

2. Notations and Preliminaries

Let $N = \{1, \ldots, n\}$ be a set of n servers, each associated with its own service rate and its own customers. The incoming stream of customers to server $i \in N$ is a Poisson process with rate λ_i. Service times are exponentially distributed. When working individually, server i, $i \in N$, is capable of giving a service at a rate of at most ξ_i, $\xi_i > \lambda_i$, customers per unit of time. However, the actual service rate μ_i, $\lambda_i < \mu_i \leqslant \xi_i$, may depend on the arrival rate to the system in a way that the lower the arrival rate, the slower the server. We assume that $\lambda_i < \mu_i$ to guarantee stability. When the arrival rate gets closer to the potential service rate ξ_i, the server gets more efficient and its service rate gets closer to ξ_i. In our model we assume that the actual service rate μ_i is given by the weighted geometric mean of the two parameters λ_i and ξ_i, i.e., $\mu_i = \lambda_i^{1-\alpha}\xi_i^{\alpha}$ for a given α, $0 < \alpha \leqslant 1$. For $\alpha = 1$, we have $\mu_i = \xi_i$; this is the case where the service rate is independent of the arrival rate. For $\alpha = 0.5$, the actual service rate is the geometric mean of λ_i and ξ_i, namely, $\mu_i = \sqrt{\lambda_i \xi_i}$. In general, if the maximum capacity of the server is more dominant in determining μ_i than is λ_i, then $\alpha > 0.5$, and the other way around in the opposite case.

For a given α, we let $\rho_i = \lambda_i/\mu_i = (\lambda_i/\xi_i)^\alpha$ and refer to it as the *utilization level* of server i, $i \in N$. We assume without loss of generality that the sequence ρ_i, or equivalently, the sequence λ_i/ξ_i for $i \in N$, is nonincreasing in i, and for convenience we define $\rho_0 = 1$ and $\rho_{n+1} = 0$, although we do not add servers 0 and $n+1$ to the set of servers. Finally, the quality of any service system is measured by its expected number of customers under steady-state conditions.

Suppose that a group of servers S, $\varnothing \subseteq S \subseteq N$, forms a coalition. The incoming stream of customers is Poisson with rate $\lambda(S) = \sum_{i\in S} \lambda_i$. The group provides an exponentially distributed service by a (now single) combined server whose potential service rate is $\xi(S) = \sum_{i\in S} \xi_i$. The actual service time of the servers in the group is exponential with rate $\mu(S) = \lambda(S)^{1-\alpha}\xi(S)^\alpha$. For $\alpha = 1$, $\mu(S) = \sum_{i\in S} \xi_i$.

For any $\varnothing \subset S \subseteq N$, let $V(S)$ be the expected number of customers in such a system in steady state. Define $\rho(S) = \lambda(S)/\mu(S) = (\lambda(S)/\xi(S))^\alpha$, which implies that $V(S) = \rho(S)/(1-\rho(S)) = \lambda(S)/(\mu(S)-\lambda(S)) = \lambda(S)^\alpha/(\xi(S)^\alpha - \lambda(S)^\alpha)$. In particular, $V(\{i\}) = \lambda_i/(\mu_i - \lambda_i) = \lambda_i^\alpha/(\xi_i^\alpha - \lambda_i^\alpha)$. $V(\varnothing)$ is naturally defined as zero. For technical reasons, we also let, below, $V(\{0\}) = \infty$ and $V(\{n+1\}) = 0$.

Because $V(\varnothing) = 0$ and $V(S)$, $\varnothing \subseteq S \subseteq N$, is a real number, the pair (N, V) defines a *cooperative game with transferable utility*. In particular, $V(\cdot)$ is its *characteristic function*.

Observation 1. For any $1 \leqslant i < j \leqslant n$, $V(\{i\}) \geqslant V(\{j\})$. Moreover, $\rho_i > \rho_j$ (or alternatively, $\lambda_i/\xi_i > \lambda_j/\xi_j$), if and only if $V(\{i\}) > V(\{j\})$.

Observation 1 implies that the servers are indexed in a nonincreasing order of their utilization levels, $V(\{1\}) \geqslant V(\{2\}) \geqslant \cdots \geqslant V(\{n\})$.

Next, we present two properties regarding ratios of positive real numbers.

Observation 2. $a_i/b_i \geqslant a_j/b_j$ if and only if $a_i/b_i \geqslant (a_i+a_j)/(b_i+b_j) \geqslant a_j/b_j$. Moreover, $a_i/b_i > a_j/b_j$, if and only if $a_i/b_i > (a_i+a_j)/(b_i+b_j) > a_j/b_j$.

Observation 3. Suppose that $a_i/b_i > a_j/b_j$, $a_i < a_j$, and $b_i < b_j$. Then, $a_j/b_j > (a_j - a_i)/(b_j - b_i)$.

Proof. Aiming for a contradiction, assume that $a_j/b_j \leqslant (a_j - a_i)/(b_j - b_i)$. This implies that $a_j b_i \geqslant a_i b_j$, or equivalently, that $a_j/b_j \geqslant a_i/b_i$, contradicting our assumption. □

Lemma 1. *For any two subsets* S, $T \subset N$ *such that* $S \cap T = \varnothing$ *and* $V(S) \leqslant V(T)$, $V(S) \leqslant V(S \cup T) \leqslant V(T)$. *Moreover, if* $V(S) < V(T)$, *then* $V(S) < V(S\cup T) < V(T)$.

Proof. According to the definition of $V(S)$ and Observation 1, $V(S)$ is strictly increasing in $\rho(S)$, and because $\rho(S) = (\lambda(S)/\xi(S))^\alpha$, $V(S)$ is also strictly increasing in $\lambda(S)/\xi(S)$. The proof is then concluded by using Observation 2 on the last ratio. □

Lemma 1 implies that for any pair of coalitions S and T with $S \cap T = \varnothing$, $V(S) + V(T) \geqslant V(S \cup T)$. In other words, the cooperative game we deal with is *subadditive*. In particular, for any S, $\varnothing \subseteq S \subseteq N$, $V(S) + V(N\backslash S) \geqslant V(N)$. Thus, the formation of the grand coalition, i.e., when all servers pool their arrival streams and their service capacities, is a natural outcome from a bargaining process. The lemma also shows that the total cost of a coalition that is formed by the union of two disjoint sets is somewhere in-between the costs of the two respective sets. Observe also that $V(\cdot)$ is not monotone, and if $S \subset T$, all three orders between $V(S)$ and $V(T)$ (namely, $V(S) < V(T)$, $V(S) > V(T)$, or $V(S) = V(T)$) are a priori possible.

In this article we pose the question of how to allocate the cost $V(N)$ among various servers in N so that no subset of servers would have any incentive to deviate from the grand coalition, assuming that servers indeed merge in the abovementioned way. In other words, we look for *core allocations* for the game (N, V). By that we mean that we look for vectors $x \in R^n$ such that $\sum_{i\in N} x_i = V(N)$ (this is the *cost sharing* (or Pareto-efficient) requirement) and such that for any set of servers S with $\varnothing \subseteq S \subset N$, $\sum_{i\in S} x_i \leqslant V(S)$ (this is the *stand-alone* requirement). A game is called *balanced* if its core is nonempty, and it is called *totally balanced* if all the games with the same characteristic function, but restricted to subsets of players, are balanced too.

One of the desired properties when considering the cost allocation according to a certain characteristic function is whether or not this function is concave (see the definition below). A game defined by a concave characteristic function is called a *concave game*. It is well known that concave games are totally balanced, as first shown in Shapley (1971).

Definition 1. A set function $f(\cdot)$ is said to be *concave* if for any two subsets $\varnothing \subseteq S \subset T \subset N$ and any $l \in N\backslash T$,

$$f(S \cup \{l\}) - f(S) \geqslant f(T \cup \{l\}) - f(T). \tag{1}$$

There are a few other possible (equivalent, of course) definitions for concave functions. One of them is the following.

Definition 2. A set function $f(\cdot)$ is said to be *concave* if for any two subsets $\varnothing \subseteq S$, $T \subseteq N$,

$$f(S \cup T) + f(S \cap T) \leqslant f(S) + f(T).$$

The concavity property of set functions is sometimes referred to as *submodularity*. As it turns out, our $V(S)$ is not concave. This is shown via the following example.

EXAMPLE. Consider the case where $\alpha = 1$, a set N with three servers, $N = \{1, 2, 3\}$, and the game (N, V) with the characteristic function $V(\cdot)$ as defined above. Note that for $\alpha = 1$, $\xi_i = \mu_i$ for all i, and $\xi(S) = \mu(S) = \sum_{i \in S} \xi_i$. Let

$\lambda_1 = 9$, $\xi_1 = 10$, and hence $\rho_1 = 0.9$ and $V(\{1\}) = 9$;

$\lambda_2 = 5$, $\xi_2 = 10$, and hence $\rho_2 = 0.5$ and $V(\{2\}) = 1$; and

$\lambda_3 = 1$, $\xi_3 = 10$, and hence $\rho_3 = 0.1$ and $V(\{3\}) = 1/9$.

Take $S = \{1\}$, $T = \{1, 2\}$, and $l = 3$. Then, $\lambda(\{1, 3\}) = 10$, $\xi(\{1, 3\}) = 20$, $\rho(\{1, 3\}) = 0.5$, $V(\{1, 3\}) = 1$, and $V(\{1, 3\}) - V(\{1\}) = -8$.

Also, $\lambda(\{1, 2\}) = 14$, $\xi(\{1, 2\}) = 20$, $\rho(\{1, 2\}) = 7/10$, $V(\{1, 2\}) = 7/3$; and, $\lambda(\{1, 2, 3\}) = 15$, $\xi(\{1, 2, 3\}) = 30$, $\rho(\{1, 2, 3\}) = 0.5$, $V(\{1, 2, 3\}) = 1$. Thus, $V(\{1, 2, 3\}) - V(\{1, 2\}) = 1 - 7/3 = -4/3$, meaning that inequality (1) does not hold for this choice for α, S, T, and l. However, the core of this game is nonempty. Examples for core allocations are $x_1 = 1$ with $x_2 = x_3 = 0$, or $x'_1 = \frac{36}{63}$, $x'_2 = \frac{20}{63}$ with $x'_3 = \frac{7}{63}$. Any convex combination of these two allocations is also in the core.

Of course, the fact that $V(S)$ is not concave does not imply that the core of the game is empty. As exemplified above, concavity is only a sufficient (not a necessary) condition for balance. In fact, for the special case where $\alpha = 1$, $y_i = \lambda_i/(\xi(N) - \lambda(N))$, $1 \leqslant i \leqslant n$, is a core allocation of the game (N, V), and therefore we conclude that the game (N, V) is balanced. However, for the general case of α, the balancedness of the game (N, V) still needs to be resolved. For $\alpha = 1$, because $1/(\mu(N) - \lambda(N))$ is the mean time in the system for each of the customers when the grand coalition is formed, the proposed allocation is an allocation in which each server pays the waiting costs associated with its own customers under the most efficient social use of the servers. At first sight, this allocation seems quite fair. However, it ignores the relative efficiency of the servers. For example, consider $\alpha = 1$ and two servers, with $\lambda_1 = \lambda_2 = 9$, $\xi_1 = 10$, and $\xi_2 = 100$. As we see, the two servers have the same rate of incoming customers, but the second server is 10 times more efficient than the first. However, the core cost allocation mentioned above assigns exactly the same cost to the two servers, namely, $y_1 = y_2 = \frac{9}{92}$, an allocation that does not reflect the efficiency of the second server. Indeed, the second server gains very little from this cooperation; instead of paying $V(\{2\}) = \frac{9}{91}$, she now pays $y_2 = \frac{9}{92}$, where the first server reduced his costs from $V(\{1\}) = 9$ without cooperation, to $y_1 = \frac{9}{92}$—i.e., a reduction of a factor of 92.

In the sequel, we continue and investigate the core of the game, and we propose an infinite family of core cost allocations. In particular, our methodology will result in a core cost allocation for the above example $x_1 = \frac{18}{92}$, and $x_2 = 0$, which is more sound, because the first server saves almost all of his costs by joining the second server. Hence, he can be generous and pay the total cost inflicted by the grand coalition so that the two servers are happy to cooperate. In addition, our analysis will also lead to the core allocation $x_1 = 9$ and $x_2 = -8\frac{74}{92}$. According to this allocation, the first server pays its own cost of 9 as if he worked individually, where $\frac{18}{92}$ out of 9 is used to cover the cost of the grand coalition, and all the surplus, namely $8\frac{74}{92}$, is paid to the second server. This can be looked at as payment made to make him join the grand coalition.

To continue investigating the core of the game (N, V), we distinguish between core allocations in which no server is paid by the other servers to make him/her join the grand coalition, i.e., nonnegative cost allocations, and core allocations in which at least one server is paid by the other servers, meaning cost allocations with at least one negative entry. We study the nonnegative subset of the core in §3, and the subset of the core containing cost allocations with at least one negative entry in §4.

3. The Nonnegative Core Allocations

In this section, we investigate and characterize all the nonnegative core allocations of the game (N, V), which was shown to be nonconcave in §2. For this reason, we construct a concave (and hence balanced) game (N, W) whose characteristic function $W(S)$ satisfies $W(S) \leqslant V(S)$ for any S, $\varnothing \subseteq S \subseteq N$, and $W(N) = V(N)$. Hence, any core allocation for the game (N, W) is also a core allocation for the game (N, V) (but not the other way around). This, of course, implies the balancedness of the game (N, V). A well-known result about concave games (see Shapley 1971) enables us to specify explicitly the extreme points of the core of (N, W). We then show that the core of (N, W) coincides with the nonnegative subset of the core of (N, V). Finally, we show that except for the trivial case of having a single server, i.e., $n = 1$, the nonnegative subset of the core of (N, V) contains an infinite number of allocations.

The characteristic function $W(S)$ for the auxiliary game (N, W) that we propose is

$$W(S) = \min_{T \mid S \subseteq T \subseteq N} V(T), \quad \varnothing \subseteq S \subseteq N.$$

In other words, $W(S)$ is the cost of the best coalition that the servers in S can be part of; i.e., any coalition T, which contains all servers in S, $T \supseteq S$, costs at least $W(S)$, and there exists at least one coalition $T \supseteq S$ for which $V(T) = W(S)$. We should note that the set of $\arg\min\{V(T) \mid S \subseteq T \subseteq N\}$ where $V(T) = \lambda(T)^\alpha/(\xi(T)^\alpha - \lambda(T)^\alpha) = 1/(((\xi(T) - \lambda(T))/\lambda(T) + 1)^\alpha - 1)$ is independent of α, for $\alpha \in (0, 1]$, because $V(T)$ is a strictly increasing function of $\lambda(T)/(\xi(T) - \lambda(T))$. Note, in addition, that $W(\varnothing) = 0$ and that $W(N) = V(N)$, making (N, W) a bonafide transferable utility game with the same set of players as the game (N, V), and with the same value $W(N) = V(N)$ to be shared among its players. Also, observe that the set function $W(\cdot)$ is nondecreasing, namely, $W(S \cup \{l\}) \geqslant W(S)$ for all $l \in N \backslash S$. However, the subadditivity of $W(S)$ is not immediate; it will follow from the concavity proof of $W(S)$ that we provide later. We demonstrate the calculation of $W(S)$ by using the three-player game with $\alpha = 1$ from the previous example.

Example (Continued). $W(\{1\}) = V(\{1,3\}) = V(\{1,2,3\}) = 1$, $W(\{2\}) = V(\{2,3\}) = 3/7$, $W(\{3\}) = V(3) = 1/9$, $W(\{1,2\}) = V(\{1,2,3\}) = 1$, $W(\{1,3\} = V(\{1,3\}) = V(\{1,2,3\} = 1$, $W(\{2,3\}) = V(\{2,3\}) = 3/7$ and $W(\{1,2,3\}) = V(\{1,2,3\}) = 1$.

We now present the two main results of this section in Theorems 1 and 2. We start with proving the concavity of the auxiliary game.

Theorem 1. 1. *The game (N, W) is concave. In particular, the game (N, W) is balanced.*
2. *The core of the game (N, V) contains the core of (N, W). In particular, the game (N, V) is balanced.*
3. *The core of the game (N, W) coincides with all nonnegative core allocations of the game (N, V).*

Proof. 1. The concavity proof of the game (N, W) for $\alpha = 1$ is deferred to the appendix. Note that at the end of this section we develop a linear-time algorithm for computing the characteristic function $W(\cdot)$ of the game (N, W). The properties of this algorithm are used in the proof. We will show here that the concavity of (N, W) for $\alpha = 1$ implies the concavity of the game (N, W) for any $\alpha \in (0, 1]$. For this sake, denote the cost $W(S)$ for any given α by

$$\Phi_\alpha(S) = \min_{T \mid S \subseteq T \subseteq N} \frac{\lambda(T)^\alpha}{\xi(T)^\alpha - \lambda(T)^\alpha}, \quad \varnothing \subseteq S \subseteq N.$$

Thus, we need to show that the set function $\Phi_\alpha(\cdot)$ is a concave set function, given that the set function $\Phi_1(\cdot)$ is concave. This is shown by noting that (1) $\Phi_\alpha(S) = 1/((\Phi_1(S)^{-1}+1)^\alpha - 1)$, (2) the real function g_α: $\Re^+ \to \Re^+$, defined by $g_\alpha(x) = 1/((x^{-1}+1)^\alpha - 1)$ is strictly increasing and strictly concave for $x > 0$ (it can be checked by verifying that $(d/dx)g_\alpha(x) > 0$ and $(d^2/dx^2)g_\alpha(x) < 0$ for $x > 0$), and (3) $\Phi_\alpha(S) = g_\alpha(\Phi_1(S))$. Then, to conclude the proof, we need to show that for any two subsets of N, S, and T, the following inequality holds: $\Phi_\alpha(S \cap T) + \Phi_\alpha(S \cup T) \leqslant \Phi_\alpha(S) + \Phi_\alpha(T)$. Towards this end, assume without loss of generality that $\Phi_1(S) \leqslant \Phi_1(T)$; thus, by the monotonicity of $\Phi_1(S)$ we have that $\Phi_1(S \cap T) \leqslant \Phi_1(S) \leqslant \Phi_1(T) \leqslant \Phi_1(S \cup T)$, and by its concavity we have that $\Phi_1(S \cap T) + \Phi_1(S \cup T) \leqslant \Phi_1(S) + \Phi_1(T)$. Let $a = \Phi_1(S \cap T)$, $b = \Phi_1(S)$, $c = \Phi_1(T)$, $d = \Phi_1(S \cup T)$, and $e = a + d - b$. Monotonicity of $\Phi_1(\cdot)$ implies that $a \leqslant b \leqslant c \leqslant d$, and with the concavity of $\Phi_1(\cdot)$ we get that $a \leqslant e \leqslant c$. Consider the following problem:

$$\begin{aligned} \min \quad & g_\alpha(x_1) + g_\alpha(x_2) \\ \text{s.t.} \quad & x_i \geqslant a \quad \text{for } i = 1, 2, \\ & x_i \leqslant d \quad \text{for } i = 1, 2, \\ & x_1 + x_2 = a + d. \end{aligned}$$

Both $(x_1, x_2) = (a, d)$ and $(x_1, x_2) = (b, e)$ are feasible solutions to this optimization problem; however, the concavity of $g_\alpha(\cdot)$ implies that the minimum is reached at an extreme point of the feasible region, i.e., at (a, d). This implies that $\Phi_\alpha(S \cap T) + \Phi_\alpha(S \cup T) \leqslant \Phi_\alpha(S) + \Phi_\alpha(T)$.

2. This item follows immediately from the above-mentioned facts that $W(S) \leqslant V(S)$, $\varnothing \subseteq S \subset N$, and $W(N) = V(N)$.

3. Finally, to prove the third item, suppose that there exists a nonnegative core allocation of V that is not in the core of W, i.e., there exists $(x_1, \ldots, x_n) \in \Re^n$ in the core of V, $x_i \geqslant 0$ for $1 \leqslant i \leqslant n$, and a coalition $S \subset N$ with $W(S) < \sum_{i \in S} x_i \leqslant V(S)$. However, as for some $\tilde{S} \supseteq S$, $W(S) = V(\tilde{S}) \geqslant \sum_{i \in \tilde{S}} x_i$ (where the last inequality follows from the fact that the vector x is in the core of V), we obtain that $\sum_{i \in S} x_i > \sum_{i \in \tilde{S}} x_i$, implying that $\sum_{i \in \tilde{S} \setminus S} x_i < 0$. This means that the vector x contains negative entries, contradicting our assumption. □

An immediate consequence of the third item of Theorem 1 is that all core allocations for the game (N, W) are nonnegative. It is possible to see that $x_1 = 4/7$, $x_2 = 1$, and $x_3 = -4/7$ is a core allocation for the game (N, V) appearing in the above example. Therefore, it is possible that the cores of the games (N, V) and (N, W) do not coincide. Put differently, it is possible that the former strictly contains the latter. Indeed, we will show in §4 that with the exception of some trivial cases, the core of (N, W) is a strict subset of the core of (N, V).

Next, we state all core allocations of the game (N, W). This, in turn, implies an explicit expression for the convex subset of the core of the game (N, V) composed of its nonnegative allocations.

Theorem 2. *Let $\pi = (\pi_1, \ldots, \pi_n)$ be a permutation of the n servers. Define the marginal contribution vector*

$$1. \quad x_{\pi_i} = W(\{\pi_1, \pi_2, \ldots, \pi_i\}) - W(\{\pi_1, \pi_2, \ldots, \pi_{i-1}\}), \quad 1 \leqslant i \leqslant n.$$

Then, x_π is a core allocation of the game (N, W) (and hence of the game (N, V)). Moreover, an allocation is in the core of the game (N, W) if and only if it is a convex combination of the $n!$ marginal contribution vectors of the game (N, W) obtained by all permutations.

$$2. \quad y_{\pi_i} = V(\{\pi_1, \pi_2, \ldots, \pi_i\}) - V(\{\pi_1, \pi_2, \ldots, \pi_{i-1}\}), \quad 1 \leqslant i \leqslant n.$$

If y_π is a nonnegative vector, then it is an extreme point of the core of the game (N, V). In particular, for such a permutation the vector x_π coincides with the respective vector y_π.

Proof. 1. From Shapley (1971) and Isciishi (1981) (see also Moulin 1995, p. 409), one can learn that the theorem follows from the facts that (1) $W(\varnothing) = 0$, (2) $W(S) \leqslant W(T)$ for any S and T with $\varnothing \subseteq S \subseteq T \subseteq N$, and (3) $W(\cdot)$ is a concave set function.

2. If y_π is a nonnegative vector, then $V(\{\pi_1, \pi_2, \ldots, \pi_i\}) \geqslant V(\{\pi_1, \pi_2, \ldots, \pi_{i-1}\})$ for $1 \leqslant i \leqslant n$. This means that for permutation π, the addition of any server

π_j with $j > i$ to a subset $S = \{\pi_1, \ldots, \pi_i\}$ will not strictly decrease the cost of S. Thus, $W(\{\pi_1, \ldots, \pi_i\}) = V(\{\pi_1, \ldots, \pi_i\})$, implying that $x_\pi = y_\pi$. We use Theorem 1, Part 3 and the previous item of this theorem to conclude that the cost allocation y_π is an extreme point of the core of (N, V). □

Theorem 2, Part 1 says the following: Order the servers in any order, and consider the n consecutive coalitions, where each contains the i $(1 \leqslant i \leqslant n)$ first servers in the permutation. Now, charge server j $(1 \leqslant j \leqslant n)$ the difference between the cost of the smallest (among the n) coalitions that contains server j and the cost of the maximum (among the n) coalitions that do not contain server j. In other words, each server is charged the marginal contribution of adding it to the maximum coalition of which the server is not a member. Any such assignment of charges among the servers in N yields a core allocation. Moreover, the core of the game (N, W) is the convex hull of such marginal contribution vectors. We should also note that by the monotonicity of the characteristic function W, all the extreme core allocations of the game (N, W) are nonnegative, and therefore all its core allocations are also nonnegative. In other words, each server is allocated a nonnegative share of $W(N)$.

Remark. Recall that the Shapley value of the game (N, W) is defined as the simple arithmetic mean among the abovementioned $n!$ allocations. This, coupled with the concavity of the characteristic function $W(\cdot)$, guarantees that the Shapley value of the game (N, W) is in the core of this game. To find the Shapley value of the game (N, V), the corresponding marginal contribution vectors need to be computed and averaged, this time with the characteristic function $V(\cdot)$. There is no guarantee that all the $n!$ marginal contribution vectors of the game (N, V) or that the corresponding Shapley value are/is in the core of the game (N, V).

The Shapley value is the most commonly used solution for transferable utility cooperative games. We do not attempt to provide a full description of its properties here. We would, however, like to mention one of them, that is, the *balanced contribution* property. However, first we would like to introduce the following notation. For coalition S with $i \in S$, we denote by S_{-i} the coalition that is as S but without i (elsewhere it is usually denoted by $S\backslash\{i\}$). It says that if players measure their value by the Shapley value, then the marginal contribution that player i gains if player j joins coalition N_{-j} is the same as the gain for player j when player i joins coalition N_{-i}. This shows the fairness of the Shapley cost allocation. Importantly, it is the only cost allocation possessing the balanced contribution property. For more, see Osborne and Rubinstein (1994, p. 291).

Example (Continued). There are four extreme core allocations for the game (N, W). True, there are six permutations, but as it turned out there are some repetitions in marginal contribution vectors. Specifically,

Permutation	x_1	x_2	x_3
1,2,3	1	0	0
1,3,2	1	0	0
2,1,3	4/7	3/7	0
2,3,1	4/7	3/7	0
3,1,2	8/9	0	1/9
3,2,1	4/7	20/63	1/9

The Shapley value of the game (N, W) is then $x_1 = 145/189$, $x_2 = 37/189$, and $x_3 = 1/27$. As for the game (N, V), the vectors of marginal contributions are

Permutation	x_1	x_2	x_3
1,2,3	9	$-6\frac{2}{3}$	$-1\frac{1}{3}$
1,3,2	9	0	-8
2,1,3	$1\frac{1}{3}$	1	$-1\frac{1}{3}$
2,3,1	4/7	1	$-4/7$
3,1,2	8/9	0	1/9
3,2,1	4/7	20/63	1/9

It is easy to see that not all of the marginal contribution vectors of the game (N, V) are in its core. For example, the vector associated with permutation 1, 3, 2 satisfies $x_1 + x_2 = 9 > V(\{1, 2\}) = 7/3$. According to Theorem 2, Part 2, the two nonnegative allocations—namely, the two last ones—are in the core of (N, V). It is easy to check that the two middle ones are also in the core of (N, V). By averaging these six vectors, we get the Shapley value of the game (N, V): $(\frac{673}{189}, -\frac{137}{189}, -\frac{347}{189})$. This allocation is not in the core of the game (N, V) because according to this allocation the total cost allocated to Servers 1 and 2 is $536/189$, which is larger than $V(\{1, 2\}) = 7/3$.

Next, we present a constructive algorithm whose input is the game (N, V) and which returns the value of $W(S)$ for any $\varnothing \subseteq S \subseteq N$. We note that in general there may be more than one coalition $T \supseteq S$, for which $V(T) = W(S)$, as was demonstrated by the above example. The constructive algorithm returns the largest coalition $T \supseteq S$ for which $V(T) = W(S)$. The next lemma claims that such a largest coalition exists.

Lemma 2. *For any given set* $S \subset N$, *if* $T_1, T_2 \in \arg\min_{T \mid S \subseteq T \subseteq N} V(T)$ *with* $T_1 \neq T_2$, *then* $T_1 \cup T_2 \in \arg\min_{T \mid S \subseteq T \subseteq N} V(T)$.

Proof. In view of the fact that the set $\arg\min_{T \mid S \subseteq T \subseteq N} V(T)$ is independent of α, it is sufficient to prove the lemma for $\alpha = 1$. Thus, we assume this value for α throughout the proof. Suppose that for a given set $S \subset N$, $T_1, T_2 \in \arg\min_{T \mid S \subseteq T \subseteq N} V(T)$ with $T_1 \neq T_2$. By definition

$V(T_1) = V(T_2)$. Clearly, any server i, $i \notin S$ that satisfies $V(\{i\}) > V(T_1)$ is not a member of either T_1 or T_2 because otherwise, by Observation 2, its removal would have strictly decreased the cost of the set, contradicting the definition of T_1 and T_2. Moreover, any server i, $i \notin S$ that satisfies $V(\{i\}) < V(T_1)$ is a member of T_1 and of T_2 because otherwise, by Observation 2, its deletion would have strictly increased the cost of the set, contradicting the definition of T_1 and T_2. Thus, T_1 and T_2 may differ only with respect to servers i, $i \notin S$ for which $V(\{i\}) = V(T_1) = V(T_2)$. By Observation 2, it follows that adding such a server to either T_1 or T_2 does not affect the cost of the sets. Therefore, $T_1 \cup T_2 \in \arg\min_{T \mid S \subseteq T \subseteq N} V(T)$. □

In view of Lemma 2, we are now ready for the next definition.

Definition 3. The optimal coalition for S, $\varnothing \subseteq S \subseteq N$, is the largest coalition T for which $W(S) = V(T)$ and $S \subseteq T \subseteq N$. We denote the optimal coalition for S by $\widetilde{S}$.

The constructive algorithm that we propose below returns for any S, $S \subseteq N$, in linear time in n, both $W(S)$ and the *optimal coalition* for S, namely, $\widetilde{S}$. The algorithm is a greedy one. Specifically, it adds to S servers from $N\backslash S$ one by one, commencing with the one with the highest index there, then with the second highest, etc. This continues as long as by doing so the value of the coalition does not increase. We next bring the proof of the correctness of the algorithm.

The Construction Algorithm Returning the Optimal Coalition of S:
input: S, $\varnothing \subseteq S \subset N$.
output: $W(S)$ and $\widetilde{S}$.
Step 0: Let $\Phi_1(S) = \lambda(S)/(\xi(S) - \lambda(S))$ (the expected cost when $\alpha = 1$), $\widetilde{S} = S$; $S' = N\backslash S$; Let $c_S = \max S'$ and $b_S = n + 1$.
Step 1: While $\Phi_1(\widetilde{S}) \geqslant \Phi_1(\{c_S\})$ let $\widetilde{S} \leftarrow \widetilde{S} \cup \{c_S\}$; $b_S = c_S$; $S' \leftarrow S'\backslash\{c_S\}$; If $S' \neq \varnothing$ then let $c_S = \max S'$. Otherwise, let $c_S = 0$ and goto Step 2. Endwhile.
Step 2: Let $W(S) = V(\widetilde{S})$. Return $\widetilde{S}$, $W(S)$, b_S, and c_S.
End of Algorithm.

Remark. Note that the terminal value for b_S is the smallest index added to S towards the formation of $\widetilde{S}$, whereas c_S is the highest index left out.

Theorem 3. *The coalition $\widetilde{S}$ obtained by the Construction Algorithm is the optimal coalition for S.*

Proof. Note that by Observation 2 and the definition of the algorithm, $V(\{b_S\}) \leqslant V(\widetilde{S}) < V(\{c_S\})$. The same observation implies that inserting to the coalition $\widetilde{S}$ any subset of servers $\Psi \subset N\backslash\widetilde{S}$ will strictly increase the cost of the coalition because $V(\Psi) \geqslant V(\{c_S\}) > V(\widetilde{S})$. Moreover, removing from the coalition $\widetilde{S}$ any subset $\Upsilon \subset \widetilde{S}\backslash S$ will either increase the cost of the coalition formed by the remaining servers or it will leave it unchanged, i.e., $V(\widetilde{S}\backslash\Upsilon) \geqslant V(\widetilde{S})$. Thus, the terminating $\widetilde{S}$ is indeed the optimal coalition for S. □

Example (Continued). Note that $\widetilde{\{1\}} = \{1, 2, 3\}$, although $V(\{1, 3\}) = V(\{1, 2, 3\})$ as the algorithm returns the largest best coalition. $\widetilde{\{2\}} = \{2, 3\}$, $\widetilde{\{3\}} = \{3\}$, $\widetilde{\{1, 2\}} = \{1, 2, 3\}$, $\widetilde{\{1, 3\}} = \{1, 2, 3\}$, $\widetilde{\{2, 3\}} = \{2, 3\}$, and $\widetilde{\{1, 2, 3\}} = \{1, 2, 3\}$.

For the pathological case where $n = 1$, it is clear that there exists only one core allocation for the game $(\{1\}, V)$, namely, $\lambda_1^\alpha/(\xi_1^\alpha - \lambda_1^\alpha)$ is allocated to the single server. The next theorem shows that this is the only case where the core is a singleton.

Theorem 4. *If $n \geqslant 2$, the core of the game (N, V) contains infinitely many nonnegative cost allocations.*

Proof. We next state two different core allocations for the game (N, V) when $n \geqslant 2$. These, coupled with the convexity of the core, complete the proof. First, recall the original order in which servers are indexed in a nonincreasing order of ρ_i, $i = 1, \ldots, n$. For this specific permutation $(1, 2, \ldots, n)$, one can see by Theorem 2 that the marginal contribution vector $y_{\pi_i} = W(\{1, 2, \ldots, i\}) - W(\{1, 2, \ldots, i-1\})$, $1 \leqslant i \leqslant n$, is a core allocation for both games, (N, W) and (N, V). It is easy to see that because server n is a member of the optimal coalition of any subset $S \subseteq N$, and in particular $n \in \widetilde{\{1 \ldots i\}}$ for any $i \in \{1, 2, \ldots, n\}$, $y_n = \lambda_n^\alpha/(\xi_n^\alpha - \lambda_n^\alpha)$ if $n = 1$, and $y_n = W(N) - W(N_{-i}\}) = 0$ if $n \geqslant 2$.

Consider now the reversed permutation $\pi_j = n - j + 1$, $1 \leqslant j \leqslant n$, i.e., the permutation $(n, n-1, \ldots, 1)$. Because $\{\pi_1, \pi_2, \ldots, \pi_i\} = \{n, n-1, \ldots, n-i+1\}$, by Observation 2, $V(\{\pi_1, \pi_2, \ldots, \pi_i\}) = W(\{\pi_1, \pi_2, \ldots, \pi_i\})$ because for any i, $1 \leqslant i \leqslant n$, $\{\pi_1, \pi_2, \ldots, \pi_i\}$ is the best coalition of itself (although possibly it is not the maximal best one, i.e., the optimal one). One can see by Theorem 2 that the marginal contribution vector $y'_{\pi_i} = V(\{\pi_1, \pi_2, \ldots, \pi_i\}) - V(\{\pi_1, \pi_2, \ldots, \pi_{i-1}\})$, $1 \leqslant i \leqslant n$, is a core allocation for both (N, W) and (N, V). Note that $y'_{\pi_1} = y'_n = V(\{n\}) = \lambda_n^\alpha/(\xi_n^\alpha - \lambda_n^\alpha)$. Clearly, $y'_n = y_n$ if and only if $n = 1$. □

4. Core Allocations with Payments to Servers

In the previous section we have characterized all the nonnegative core allocations of the game (N, V). Next, we investigate the possibility of core allocations, with some entries being negative. We show that, with one exception (it is when all utilization levels coincide), such allocations always exist. Note that a negative entry means that the corresponding server is paid by the others to persuade him/her to join the grand coalition. In spite of the fact that the servers that are paid have their own customers contributing to the total congestion, such payments should not be ruled out because the contribution of fast servers can outweigh these waiting costs.

We start by presenting a procedure that generates core cost allocations for (N, V) with a single negative entry.

Towards this end, let $BN = \{i\colon V(N) < V(N_{-i})\}$. In words, BN consists of the (best) servers in N in the sense that if any one of them (or by Observation 2, any subset of them) is excluded from the grand coalition, the cost of the remaining set strictly increases. Note that with the exception of the case where $V(\{i\}) = c$ for $i = 1, \ldots, n$, namely, when all utilization levels across all serves are identical, the last server, server n, is always a member of BN.

The next theorem states that those servers who are not paid by the others in all core allocations are exactly the ones whose marginal contribution to all the other servers towards the formation of the grand coalition is not strictly negative. In particular, if the set $BN = \varnothing$, or equivalently if all utilization factors are identical, then there does not exist any core allocation for the game (N, V) with negative entries.

Theorem 5. *If $i \notin BN$, then a core allocation $(x_1, \ldots, x_n)$ for (N, V) with $x_i < 0$ does not exist. In particular, if all utilization levels are identical, there does not exist any core allocation for the game (N, V) with negative entries.*

Proof. Suppose that $i \notin BN$, i.e., $V(N) \geqslant V(N_{-i})$. Let x be a core allocation of (N, V), implying that $\sum_{j=1}^{n} x_j = V(N)$ and $\sum_{j\neq i} x_j \leqslant V(N_{-i})$. Thus, $x_i \geqslant 0$. Suppose now that all servers have the same utilization levels, which is equivalent to assuming that $V(\{i\}) = c$ for $i = 1, \ldots, n$ and some $c > 0$. In this case, for any subset of servers $S \subseteq N$, $V(S) = c$, and in particular, $V(N) = V(N_{-i}) = c$ for any $i = 1, \ldots, n$, implying that $BN = \varnothing$. This concludes the proof. □

In the sequel we will present a procedure (see Theorem 6 below) that generates for each $i \in BN$ infinitely many core cost allocations for (N, V), for which server i is paid by the servers in N_{-i}. To develop these core allocations, define for any $i \in BN$ and any $S \subseteq N_{-i}$: First, $W_i(S) = \min\{V(S')\colon S \subseteq S' \subseteq N_{-i}\}$, which is the cost of the optimal coalition for S over the set of servers N_{-i}. Clearly, $W_i(S) \geqslant W(S)$ because in $W(S)$ we also allow addition of server i to the optimal coalition of S if it helps. Second, $\Omega_i(S) = \min\{W_i(S), W(S) + V(N_{-i}) - V(N)\}$. Note that $V(N_{-i}) - V(N) > 0$ because $i \in BN$, implying that $W(S) + V(N_{-i}) - V(N) > W(S)$. In addition, observe that because $i \notin S$, then $V(S) \geqslant W_i(S)$, implying that $W(S) \leqslant \Omega_i(S) \leqslant V(S)$. Thus, with the aid of the Ω_i functions, $i \in BN$, we construct tighter lower bounds on $V(S)$ for $\varnothing \subseteq S \subseteq N_i$ for $i \in BN$.

Example (Continued). Because $V(\{1,2,3\}) = 1$, $V(\{1,3\}) = 1$, and $V(\{1,2\}) = 7/3$, the only server that strictly reduces the cost when added to the other servers is server 3, implying that $BN = \{3\}$. Implementing the above notation, $N_{-3} = \{1,2\}$ and $V(N_{-3}) - V(N) = 7/3 - 1 = 4/3$. We now calculate $\Omega_3(S)$ for any $S \subseteq N_{-3}$. For $S = \{1\}$, $W_3(\{1\}) = \min\{V(S')\colon S' \subseteq N_{-3}\} = V(\{1,2\}) = 7/3$ and $W(\{1\}) = V(\{1,2,3\}) = 1$; thus, $\Omega_3(\{1\}) = \min\{7/3, 1 + 4/3\} = 7/3$. Observe the improved tightness of the lower bound for $V(\{1\})$ because $1 = W(\{1\}) < \Omega_3(\{1\}) = 7/3 < V(\{1\}) = 9$. For $S = \{2\}$, $W_3(\{2\}) = V(\{2\}) = 1$ and $W(\{2\}) + V(N_{-3}) - V(N) = V(\{2,3\}) + 4/3 = 3/7 + 4/3 = 37/21$, implying that $\Omega_3(\{2\}) = \min\{1, 37/21\} = 1$. Also, here we get an improved bound because $3/7 = W(\{2\}) < \Omega_3(\{2\}) = 1 = V(\{2\})$. Finally, for $S = N_{-3} = \{1,2\}$, $W_3(\{1,2\}) = V(\{1,2\}) = 7/3$ and $W(\{1,2\}) + V(N_{-3}) - V(N) = V(\{1,2,3\}) + 4/3 = 7/3$, implying that $\Omega_3(\{1,2\}) = \min\{7/3, 7/3\} = 7/3$. The bound is also improved in this case because $1 = W(\{1,2\}) < \Omega_3(\{1,2\}) = 7/3 = V(\{1,2\})$. In this example, for all $S \subseteq N_{-3}$, $W(S) < \Omega_3(S)$. In fact, it is easy to see that for any N with nonidentical servers with respect to their utilization levels, $W(S) < \Omega_n(S)$ for any set $S \subseteq N_{-n}$.

The set function $\Omega_i(\cdot)$ defined over subsets of N_{-i} is the minimum over two nonnegative, nondecreasing, and concave set functions, $W_i(\cdot)$ and $W(\cdot) + V(N_{-i}) - V(N)$. Therefore, $\Omega_i(\cdot)$ is nonnegative and nondecreasing in N_{-i}. In Theorem 6 Part 2, we prove that the game (N_{-i}, Ω_i) is balanced, meaning that its core is nonempty. Part 1 of Theorem 6 proves that server $i \in BN$ is never being paid more than $V(N_{-i}) - V(N)$. As shown constructively in Part 3, this bound is tight. In other words, Part 3 generates for each $i \in BN$ a core allocation for the game (N, V) in which server i is paid by the others the maximum possible, namely, $V(N_{-i}) - V(N)$. We conclude with Part 4 of the theorem, in which we show that there are infinitely many core allocations in which server i is paid by the other servers.

Theorem 6. *For any $i \in BN$,*

1. *There does not exist a core allocation for (N, V) in which the payment to server i exceeds $V(N_{-i}) - V(N)$. In other words, for any (N, V) core allocation x, $x_i \geqslant V(N) - V(N_{-i})$.*
2. *The game (N_{-i}, Ω_i) is balanced. In particular, the allocation $x_j = \lambda_j/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^{\alpha} - \lambda(N_{-i}))$ for $j \in N_{-i}$ is in its core.*
3. *The allocation y given by $y_i = V(N) - V(N_{-i})$, and $y_j = x_j$ for $j \in N_{-i}$, where x_j is the core allocation given in Part 2 above for the game (N_{-i}, Ω_i), is a core allocation of the game (N, V) having a single negative entry y_i.*
4. *There exist infinitely many core cost allocations for game (N, V) in which server i is paid by the servers in N_{-i}.*

Proof. 1. Let x be a core cost allocation for the game (N, V), implying that $\sum_{j\in N} x_j = V(N)$ and $\sum_{j\in N_{-i}} x_j \leqslant V(N_{-i})$. By using these two constraints, we conclude that $x_i = V(N) - \sum_{j\in N_{-i}} x_j \geqslant V(N) - V(N_{-i})$.

2. To show that the allocation $x_j = \lambda_j/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^{\alpha} - \lambda(N_{-i}))$, $j \in N_{-i}$, is in the core of the game (N_{-i}, Ω_i), we prove that $\sum_{j\neq i} x_j = \Omega_i(N_{-i})$ and that for any $S \subset N_{-i}$, $\sum_{j\in S} x_j \leqslant \Omega_i(S)$.

First, observe that $\sum_{j\neq i} x_j = \lambda(N_{-i})/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^{\alpha} - \lambda(N_{-i})) = \lambda(N_{-i})^{\alpha}/(\xi(N_{-i})^{\alpha} - \lambda(N_{-i})^{\alpha}) = V(N_{-i}) = W_i(N_{-i}) = W(N_{-i}) + V(N_{-i}) - V(N)$, where the last equality follows from the fact that

$i \in BN$, implying that $W(N_{-i}) = V(N)$. By using the definition of Ω_i, we conclude that $\sum_{j\neq i} x_j = \Omega_i(N_{-i})$.

To prove the stand-alone inequalities, we first present the following observation:

Observation 4. Let $S \subset N$; then, the set function $\Psi(S) = \lambda(S)^{1-\alpha}\xi(S)^\alpha - \lambda(S)$ is strictly increasing in S.

To prove this observation, it suffices to show that for any set $\varnothing \neq T \subset N$ such that $S \cap T = \varnothing$, $\Psi(S \cup T) > \Psi(S)$. Let $\sum_{j\in T} \lambda_j = \Delta$. Clearly, $\sum_{j\in T} \xi_j > \Delta$. Thus, $\Psi(S\cup T) > (\lambda(S)+\Delta)^{1-\alpha}(\xi(S)+\Delta)^\alpha - (\lambda(S)+\Delta)$. It suffices to show that $(\lambda(S) + \Delta)^{1-\alpha}(\xi(S) + \Delta)^\alpha \geqslant \lambda(S)^{1-\alpha}\xi(S)^\alpha + \Delta$. To show that this last inequality holds, we show that the real function $q(\Delta) = (a+\Delta)^{1-\alpha}(b+\Delta)^\alpha - a^{1-\alpha}b^\alpha - \Delta$ for any given a, $b > 0$ is nondecreasing in Δ, and because $q(0) = 0$ it will conclude the proof. This is indeed the case because $q'(\Delta) = ((1-\alpha)(b+\Delta) + \alpha(a+\Delta) - (a+\Delta)^\alpha(b+\Delta)^{1-\alpha})/((a+\Delta)^\alpha(b+\Delta)^{1-\alpha}) \geqslant 0$. The last inequality can be argued as follows. The numerator is positive because it is the difference between a (weighted) arithmetic mean and a (weighted) geometric mean between two positive numbers. Finally, the denominator is clearly a positive number. This concludes the proof of the observation.

We continue with the proof of the stand-alone inequalities. This is done below while considering two cases: First, when $i \notin \tilde{S}$, and then when $i \in \tilde{S}$.

Case 1: $i \notin \tilde{S}$. $i \notin \tilde{S}$ implies that $S \subset N_{-i}$ and $S \neq N_{-i}$ (because $i \in \tilde{N}_{-i}$ for $i \in BN$).

$$\begin{aligned}\sum_{j\in S} x_j &= \frac{\lambda(S)}{\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})}\\ &\leqslant \frac{\lambda(\tilde{S})}{\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})}\\ &< W(S) < W(S) + (V(N_{-i}) - V(N)).\end{aligned}$$

The first strong inequality follows from Observation 4 and the fact that $\tilde{S} \subset N_{-i}$ because $i \notin BN$, which implies that $\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i}) \geqslant \lambda(\tilde{S})^{1-\alpha}\xi(\tilde{S})^\alpha - \lambda(\tilde{S})$. In addition, note that $\sum_{j\in S} x_j = \lambda(S)/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) \leqslant \lambda(\tilde{S})/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) \leqslant W_i(S)$ where the last inequality follows from Observation 4 and $i \notin \tilde{S}$. Thus, for any $S \subset N_{-i}$, $i \notin \tilde{S}$, $\sum_{j\in S} x_j \leqslant \min\{W_i(S), W(S) + (V(N_{-i}) - V(N))\} = \Omega_i(S)$.

Case 2: $i \in \tilde{S}$. For $S \subset N_{-i}$ such that $i \in \tilde{S}$, we have: $\sum_{j\in S} x_j = \lambda(S)/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) \leqslant W_i(S)$ because in $W_i(S)$ the numerator cannot be smaller than $\lambda(S)$ and by Observation 4, the denominator cannot be larger than $\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})$. It remains to show that $\sum_{j\in S} x_j \leqslant W(S) + (V(N_{-i}) - V(N))$. Towards this end, consider the following trivial equation: $\sum_{j\in S} x_j = \lambda(S)/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) = (\lambda(S) + \lambda_i)/(\lambda(N)^{1-\alpha}\xi(N)^\alpha - \lambda(N)) + [\lambda(S)/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) - (\lambda(S) + \lambda_i)/(\lambda(N)^{1-\alpha}\xi(N)^\alpha - \lambda(N))]$. The first term in the right-hand side of the equation is smaller than or equal to $W(S)$ because the numerator in $W(S)$ cannot be smaller than $\lambda(S) + \lambda_i$ in view of the fact that $i \in \tilde{S}$, and the denominator of $W(S)$ cannot be larger than $\lambda(N)^{1-\alpha}\xi(N)^\alpha - \lambda(N)$ in view of Observation 4. We now consider the expression in the square brackets. We will show that $\lambda(S)/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) - (\lambda(S)+\lambda_i)/(\lambda(N)^{1-\alpha}\xi(N)^\alpha - \lambda(N)) < V(N_{-i}) - V(N) = \lambda(N_{-i})/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) - (\lambda(N_{-i}) + \lambda_i)/(\lambda(N)^{1-\alpha}\xi(N)^\alpha - \lambda(N))$. This inequality is established by showing that $(\lambda(N_{-i}\backslash S))/(\lambda(N_{-i})^{1-\alpha}\xi(N_{-i})^\alpha - \lambda(N_{-i})) - (\lambda(N_{-i}\backslash S))/(\lambda(N)^{1-\alpha}\xi(N)^\alpha - \lambda(N)) > 0$, which holds by Observation 4. Thus, for $S \subset N_{-i}$, $i \in \tilde{S}$, we have that $\sum_{j\in S} x_j < W(S) + (V(N_{-i}) - V(N))$, implying that also for this case $\sum_{j\in S} x_j \leqslant \min\{W_i(S), W(S) + (V(N_{-i}) - V(N))\} = \Omega_i(S)$. Thus, we conclude that $(x_j)_{j\in N_{-i}}$ is in the core of the game (N_{-i}, Ω_i), proving that this game is balanced.

3. By definition, $y_i < 0$ and $y_j > 0$ for $j \in N_{-i}$. To show that this allocation is in the core of the game (N, V) we need to prove that $\sum_{j\in N} y_j = V(N)$, and that for any $S \subset N$, $S \neq N$, $\sum_{j\in S} y_j \leqslant V(S)$. First, note that $\sum_{j=1}^n y_j = y_i + \sum_{j\in N_{-i}} y_j = (V(N) - V(N_{-i})) + \Omega_i(N_{-i}) = (V(N) - V(N_{-i})) + V(N_{-i}) = V(N)$, using the fact that $\Omega_i(N_{-i}) = V(N_{-i})$. It remains to show that for any $S \subseteq N$, $\sum_{j\in S} y_j \leqslant V(S)$. Here, too, we distinguish between two cases:

Case 1: $i \notin S$. Because y_j, $j \in N_{-i}$, is a core allocation of (N_{-i}, Ω_i), we get that $\sum_{j\in S} y_j \leqslant \Omega_i(S) \leqslant W_i(S) \leqslant V(S)$.

Case 2: $i \in S$. First, $\sum_{j\in S} y_j = y_i + \sum_{j\in S\backslash\{i\}} y_j \leqslant y_i + \Omega_i(S\backslash\{i\}) \leqslant y_i + W(S\backslash\{i\}) + (V(N_{-i}) - V(N)) = W(S\backslash\{i\}) \leqslant W(S) \leqslant V(S)$, where the third inequality follows from W being a nondecreasing set function. This completes the proof that y is a core allocation of the game (N, V).

4. Let $y^1 = y$ be the core allocation of (N, V) described in Part 3, with $y_i^1 = V(N) - V(N_{-i}) < 0$ and $y_j^1 > 0$ for $j \in N_{-i}$. We now propose another core allocation y^2 for (N, V) which is nonnegative, and in addition $y_i^2 = 0$. Clearly, any (nontrivial) convex combination y' of y^1 and y^2 is a core cost allocation for (N, V) with $y_i' < 0$ and $y_j' \geqslant 0$ for $j \in N_{-i}$, proving the statement. To complete the proof and construct y^2, we invoke Theorem 2, Part 1 on a permutation π of N in which i is the last server in the permutation, i.e., $\pi_n = i$. Recall that by our assumptions, adding server i to N_{-i} strictly reduces the total cost. Therefore, it follows that $W(N_{-i}) = V(N) = W(N)$. Thus, in the respective marginal contribution vector, server i is allocated $W(N) - W(N_{-i}) = 0$, and the other servers are allocated nonnegative costs. □

We remark that as our procedure generates core allocations with a single negative entry for each $i \in BN$, convex combinations of such vectors may produce core allocations with multiple negative entries if BN consists of more than one server.

Example (Continued). Implementing Theorem 6, Part 3 in our example, we get the following core allocation

$x=(x_1, x_2, x_3)$ such that $x_3 = V(N) - V(N_{-3}) = 1-7/3 = -4/3$. $x_1 = \lambda_1/(\xi(N_{-3}) - \lambda(N_{-3})) = 9/(20-14) = 1.5$, and $x_2 = (\lambda_2/(\xi(N_{-3}) - \lambda(N_{-3})) = 5/(20-14) = 5/6$. To sum up this example, we note that in §3 we got four different nonnegative extreme core allocations for the game (N, W) by finding all marginal contribution vectors of W, namely, $y^1 = (1, 0, 0)$, $y^2 = (4/7, 3/7, 0)$, $y^3 = (8/9, 0, 1/9)$, and $y^4 = (4/7, 20/63, 1/9)$. The last two of these, namely y^3 and y^4 were also obtained as (nonnegative) marginal contribution vectors of the game (N, V), and hence by Theorem 2, Part 2 are also extreme core allocations of the game (N, V). A direct calculation of the extreme points in the core of the game (N, V) reveals that there are five such points. In addition to the abovementioned y^3 and y^4, also $(4/7, 1, -4/7)$, $(4/3, 1, -4/3)$, $(7/3, 0, -4/3)$ are extreme core allocations of the game (N, V). Indeed, because $BN = \{3\}$, it is only the third server who is paid by the others under all three extreme core allocations with a negative entry. In two of these last three core allocations, the third server is paid $\frac{4}{3}$, which is the maximum possible according to Theorem 6, Part 1. Recall that the core allocation $(4/7, 1, -4/7)$ was also obtained as the marginal contribution vector of the game (N, V) with respect to the permutation $(2, 3, 1)$.

5. Concluding Remarks

In this paper we defined a transferable utility cooperative game resulting when servers combine their capacities and their customer populations to form one service provider and one arrival stream. We then show that when the value of a coalition of servers is defined as the mean steady-state number of customers in the pooled system, then the resulting game is balanced. We were also able to give an explicit expression for a convex subset of its core, despite the fact that the game is neither monotone nor concave. This was done by defining an auxiliary concave game whose characteristic function is dominated by the function of the original game and shares the same value for the grand coalition. As it turned out, the core of this game coincides with the nonnegative core allocations of the original game, and hence the later subset was fully characterized. Finally, we identified a subset of core allocations with some of its entries being negative, and we identified the set of servers for which negative entries never exist.

Next, we state a few possible generalizations and variations that can be looked at in future research.

In all of our derivations above, we assumed that $\lambda_i < \xi_i$, $1 \leqslant i \leqslant n$. However, there is room for dealing with the case where only a weaker assumption is assumed, that is, $\lambda(N) < \xi(N)$. In particular, individual servers, or even some coalitions, may not be in a position to serve all their customers. For such coalitions S, it is natural to assume that $V(S) = \infty$. Hence, the formation of a coalition is first of all to be able to serve all customers. We can now pose all the questions that were posed above for this more general case. An observation we like to make here is that now the game (N, W) is not necessarily concave. As a counterexample, consider the following four-player game with $\alpha = 1$, implying that $\xi_i = \mu_i$:

$\lambda_1 = 2$, $\xi_1 = 1$ (hence, $\rho_1 = 2 > 1$), $\lambda_2 = 2$, $\xi_2 = 1$ (hence, $\rho_2 = 2 > 1$), $\lambda_3 = 1$, $\xi_3 = 2$ (hence, $\rho_3 = 1/2$), and $\lambda_4 = 1$, $\xi_4 = 3$ (hence, $\rho_4 = 1/3$). Take $S = \{4\}$, implying that $\widetilde{S} = \{4\}$ and $V(\widetilde{S}) = W(S) = 1/2$. Then, take $T = \{2, 4\}$, $\lambda(T) = 3$, $\xi(T) = 4$ (hence, $\rho(T) = 3/4$), where $\widetilde{T} = \{2, 3, 4\}$ with $\lambda(\widetilde{T}) = 4$ and $\xi(\widetilde{T}) = 6$. Hence, $V(\widetilde{T}) = W(T) = 2$. Finally, take $l = 1$. Then, $S \cup \{l\} = \{1, 4\}$, $\rho(S \cup \{l\}) = 3/4$ where $\widetilde{S \cup \{l\}} = \{1, 3, 4\}$, $W(S \cup \{l\}) = 2$. Thus,

$$W(S \cup \{l\}) - W(S) = 2 - 0.5 = 1.5.$$

Likewise, $T \cup \{l\} = \{1, 2, 4\}$, $\lambda(T \cup \{l\}) = 5$, $\xi(T \cup \{l\}) = 5$, $\widetilde{T \cup \{l\}} = N$, and $W(T \cup \{l\}) = V(N) = 6$. Thus,

$$W(T \cup \{l\}) - W(T) = 6 - 2 = 4,$$

which is larger than $W(S \cup \{l\}) - W(S)$. In particular, W is not concave.

Another future research topic can deal with different cost functions. Indeed, the cost function $v(S)$ we used here is quite natural, but it is certainly not the only option. One possible alternative is to keep the servers as they are (and hence let them work individually by their original service rates), but to route the combined arrival stream to the various queues in a sensible way—for example, in a way that minimizes the overall mean waiting time. There are a few options here. One is when customers join the shortest queue (breaking ties randomly), join the queue promising their minimal waiting time (given the queue lengths in front of all servers), or join the queue in a way that minimizes the overall mean waiting time (which is not the same as the previous individual criterion). Another option is when routing is done without observing the queue lengths upon arrival (and later jockeying is not allowed), but rather basing it on some statistical measures. There are two suboptions here. One is where customers join servers in the equilibrium way (trying to minimize their own waiting time), whereas in another there is a central planner who decides on the splitting of the combined arrival stream among the servers so as to minimize the overall mean waiting time. See Bell and Stidham (1983) for both equilibrium and social optimal splitting of the traffic among servers. Each of the abovementioned options leads to a different game among servers, and it will be an interesting future research project to check the properties of these games, such as finding which among them (if any) are balanced.

Appendix: Proof of Theorem 1, Part 1

This appendix is devoted to proving Theorem 1. We begin with a couple of properties satisfied by coalition $\widetilde{S}$ obtained by the Construction Algorithm on any subset $S \subseteq N$. These properties are used in the proof of Theorem 1.

LEMMA 3. *For any set $S \subseteq N$, and $\widetilde{S}$ obtained by the Construction Algorithm, the following hold*:

1. $V(\{c_S\}) > V(\widetilde{S}) \geqslant V(\{b_S\})$.
2. *$S \subset T$ implies that $\widetilde{S} \subseteq \widetilde{T}$, or equivalently, $b_S \geqslant b_T$. Moreover, $W(S) \leqslant W(T)$. Finally, if $\widetilde{S} \neq \widetilde{T}$, then $W(S) < W(T)$.*

PROOF. 1. If no server is added to S by the Construction Algorithm, then $b_S = n+1$, and by definition $V(\{n+1\}) = 0$, implying that $V(\widetilde{S}) = V(S) \geqslant V(\{b_S\}) = 0$. Otherwise, $\widetilde{S} \neq S$, and by the Construction Algorithm applied to S, server b_S is the last added to $\widetilde{S}$. Thus, $V(\widetilde{S} \backslash \{b_S\}) \geqslant V(\{b_S\})$, implying, by Lemma 1, that $V(\widetilde{S}) \geqslant V(\{b_S\})$. If $c_S = 0$, the proof is trivial because $V(\{c_S\}) = \infty$. Otherwise, $\widetilde{S} \neq N$, and the algorithm terminates when $V(\{c_S\}) > V(\widetilde{S})$.

2. We are given that $S \subset T$. Suppose that $i \notin T$ and $i \in \widetilde{S} \backslash S$. We need to show that $i \in \widetilde{T} \backslash T$. By definition of the Construction Algorithm on S, we get that $V(S \cup \{i, \ldots, n\}) \leqslant V(S \cup \{i+1, \ldots, n\})$. Because $i \notin T$, we also get that $V(S \cup \{i, \ldots, n\} \cup T) \leqslant V(S \cup \{i+1, \ldots, n\} \cup T)$ because we add exactly the same set of servers to both sides of the previous inequality. This means that when applying the Construction Algorithm to T, server i will be inserted also to $\widetilde{T}$, i.e., $\widetilde{S} \subset \widetilde{T}$, or equivalently, $b_S \geqslant b_T$. The rest of the claim follows from Observation 2. □

Proof of Theorem 1, Part 1

We now proceed to the concavity proof of the function $W(S)$.

PROOF. To prove that $W(\cdot)$ is concave, we refer to Definition 1 of concavity. Towards this end, let $S \subset T \subset N$ and $l \notin T$, and denote $S^l = S \cup \{l\}$ and $T^l = T \cup \{l\}$. We next show that

$$W(S^l) - W(S) \geqslant W(T^l) - W(T). \tag{2}$$

Both sides of the desired inequality are nonnegative because W is a monotone-increasing function, namely, adding a server to a subset can only increase the cost measured by the set function $W(\cdot)$. Therefore, if $l \in \widetilde{T}$, the inequality holds trivially because $\widetilde{T}^l = \widetilde{T}$ and $W(T^l) = W(T)$, making the right-hand side of (2) equal to zero. Also, if $\widetilde{T} = \widetilde{S}$, the inequality holds as an equality because it implies that $\widetilde{T}^l = \widetilde{S}^l$. Thus, we assume now that $l \notin \widetilde{T}$ and $\widetilde{T} \backslash \widetilde{S} \neq \varnothing$. Observe that because $S \subset T$ and $l \notin \widetilde{T}$, by Part 1 of Lemma 3 the following inequalities hold:

$$b_S \geqslant b_T \geqslant b_{T^l} > l$$

and

$$b_S \geqslant b_{S^l} \geqslant b_{T^l} > l.$$

Because $l \notin \widetilde{T}$, by applying the Construction Algorithm to T as its input, we get that $V(\{l\}) > V(\widetilde{T})$, and therefore $V(\{l\}) > V(\widetilde{T}^l) > V(\widetilde{T})$ because the coalition $\widetilde{T}^l$ contains at least one server (in fact, it is server l), which is strictly more expensive than the cost of the optimal coalition of T. Moreover, because $\widetilde{T} \backslash \widetilde{S} \neq \varnothing$, we get from Lemma 3, Part 2 that $V(\widetilde{T}) > V(\widetilde{S})$. Thus,

$$\frac{\lambda_l}{\mu_l - \lambda_l} > \frac{\lambda(\widetilde{T}^l)}{\mu(\widetilde{T}^l) - \lambda(\widetilde{T}^l)} > \frac{\lambda(\widetilde{T})}{\mu(\widetilde{T}) - \lambda(\widetilde{T})} > \frac{\lambda(\widetilde{S})}{\mu(\widetilde{S}) - \lambda(\widetilde{S})},$$

and by similar reasons,

$$\frac{\lambda_l}{\mu_l - \lambda_l} > \frac{\lambda(\widetilde{S}^l)}{\mu(\widetilde{S}^l) - \lambda(\widetilde{S}^l)} > \frac{\lambda(\widetilde{S})}{\mu(\widetilde{S}) - \lambda(\widetilde{S})}.$$

From our definitions it then follows that $\lambda(\widetilde{T}^l) > \lambda(\widetilde{T}) > \lambda(\widetilde{S})$, $\lambda(\widetilde{S}^l) > \lambda(\widetilde{S})$, $\mu(\widetilde{T}^l) > \mu(\widetilde{T}) > \mu(\widetilde{S})$, and $\mu(\widetilde{S}^l) > \mu(\widetilde{S})$.

In the rest of the proof we distinguish between two exclusive and mutually exhaustive cases.

Case (i): $b_{S^l} > c_T$. First, let $\widehat{S} = S \cup \{b_{S^l}, \ldots, n\} = \widetilde{S}^l \backslash \{l\}$. Observe that because $\widetilde{T} \supseteq S$ and $\widetilde{T} \supseteq \{c_T + 1, \ldots, n\}$, we get that $\widetilde{T} \supseteq \widehat{S}$, and therefore, $\lambda(\widetilde{T}) \geqslant \lambda(\widehat{S})$ and $\mu(\widetilde{T}) \geqslant \mu(\widehat{S})$. Also, note that $W(T^l) = V(\widetilde{T}^l) \leqslant V(\{l\} \cup \widetilde{T})$, and that $W(S^l) = V(\widetilde{S}^l) = V(\{l\} \cup \widehat{S})$. Also, $W(S) = V(\widetilde{S}) \leqslant V(\widehat{S})$. Hence, to prove the desired inequality (2), it is sufficient to show that

$$V(\{l\} \cup \widetilde{T}) - V(\{l\} \cup \widehat{S}) \leqslant V(\widetilde{T}) - V(\widehat{S}). \tag{3}$$

Towards this end, note that $V(\{l\} \cup \widetilde{T}) = (\lambda(\widetilde{T}) + \lambda_l)/(\mu(\widetilde{T}) + \mu_l - (\lambda(\widetilde{T}) + \lambda_l))$ and $V(\{l\} \cup \widehat{S}) = (\lambda(\widehat{S}) + \lambda_l)/(\mu(\widehat{S}) + \mu_l - (\lambda(\widehat{S}) + \lambda_l))$. Using this notation, inequality (3) is equivalent to

$$\frac{\lambda(\widetilde{T}) + \lambda_l}{\mu(\widetilde{T}) + \mu_l - (\lambda(\widetilde{T}) + \lambda_l)} - \frac{\lambda(\widehat{S}) + \lambda_l}{\mu(\widehat{S}) + \mu_l - (\lambda(\widehat{S}) + \lambda_l)} \leqslant \frac{\lambda(\widetilde{T})}{\mu(\widetilde{T}) - \lambda(\widetilde{T})} - \frac{\lambda(\widehat{S})}{\mu(\widehat{S}) - \lambda(\widehat{S})},$$

which by some algebra is equivalent to

$$\begin{aligned}
&\lambda(\widetilde{T})(\mu(\widehat{S}) + \mu_l - (\lambda(\widehat{S}) + \lambda_l))(\mu(\widehat{S}) - \lambda(\widehat{S})) \\
&\cdot(\mu_l - \lambda_l) - \lambda(\widehat{S})(\mu(\widetilde{T}) + \mu_l - (\lambda(\widetilde{T}) + \lambda_l))(\mu(\widetilde{T}) - \lambda(\widetilde{T})) \\
&\cdot(\mu_l - \lambda_l) + \lambda_l(\mu(\widetilde{T}) - \lambda(\widetilde{T}))(\mu(\widehat{S}) - \lambda(\widehat{S})) \\
&\cdot(\mu(\widetilde{T}) - \mu(\widehat{S}) - (\lambda(\widetilde{T}) - \lambda(\widehat{S}))) \geqslant 0.
\end{aligned}$$

Divide the last inequality by the positive constant $(\mu(\widehat{S}) - \lambda(\widehat{S}))(\mu(\widetilde{T}) - \lambda(\widetilde{T}))(\mu_l - \lambda_l)$. It then remains to show that

$$\begin{aligned}
&V(\widetilde{T})(\mu(\widehat{S}) + \mu_l - (\lambda(\widehat{S}) + \lambda_l)) \\
&\quad - V(\widehat{S})(\mu(\widetilde{T}) + \mu_l - (\lambda(\widetilde{T}) + \lambda_l)) \\
&\quad + V(\{l\})(\mu(\widetilde{T}) - \mu(\widehat{S}) - (\lambda(\widetilde{T}) - \lambda(\widehat{S}))) \geqslant 0.
\end{aligned}$$

However, the left-hand side of the last inequality is equal to

$$\begin{aligned}&(V(\{l\})-V(\hat{S}))(\mu(\tilde{T})+\mu_l-(\lambda(\tilde{T})+\lambda_l))\\&\quad-(V(\{l\})-V(\tilde{T}))(\mu(\hat{S})+\mu_l-(\lambda(\hat{S})+\lambda_l))\\&=(V(\{l\})-V(\hat{S}))(\mu(\hat{S})+\mu_l+(\mu(\tilde{T})-\mu(\hat{S}))\\&\qquad\qquad-(\lambda(\hat{S})+\lambda_l+(\lambda(\tilde{T})-\lambda(\hat{S}))))\\&\quad-(V(\{l\})-V(\tilde{T}))(\mu(\hat{S})+\mu_l-(\lambda(\hat{S})+\lambda_l))\\&=(V(\tilde{T})-V(\hat{S}))(\mu(\hat{S})+\mu_l-(\lambda(\hat{S})+\lambda_l))\\&\quad+(V(\{l\})-V(\hat{S}))(\mu(\tilde{T})-\mu(\hat{S})-(\lambda(\tilde{T})-\lambda(\hat{S}))).\end{aligned}$$

The inequality now follows from $\tilde{T}\supseteq\hat{S}$ and the fact that all terms in the last expression are nonnegative or strictly positive, i.e., $V(\tilde{T})-V(\hat{S})\geqslant 0$, $\mu(\hat{S})+\mu_l-(\lambda(\hat{S})+\lambda_l)>0$, $V(\{l\})-V(\hat{S})>V(\{l\})-V(\tilde{S}^l)>0$, and $\mu(\tilde{T})-\mu(\hat{S})-(\lambda(\tilde{T})-\lambda(\hat{S}))>0$.

Case (ii): $b_{S^l}\leqslant c_T$. In this case, $l<b_{T^l}\leqslant b_{S^l}\leqslant c_T<b_T\leqslant b_S$. Let $\Theta=\{1,\ldots,b_{S^l}-1\}\cap T\backslash S$. If Θ is empty then $\tilde{T}^l=\tilde{S}^l$ and inequality (3) follows trivially. Thus, we consider next the case where $\Theta\neq\varnothing$. Note that $T^l\subseteq\tilde{S}^l\cup\Theta$ and $\tilde{S}^l\cap\Theta=\varnothing$. In particular, we get that

$$W(T^l)\leqslant V(\tilde{S}^l\cup\Theta)=\frac{\lambda(\tilde{S}^l)+\lambda(\Theta)}{\mu(\tilde{S}^l)+\mu(\Theta)-(\lambda(\tilde{S}^l)+\lambda(\Theta))},$$

where the inequality follows from the fact that $W(T^l)$, which is the cost of the optimal coalition for T^l, is a lower bound for the cost of any other coalition that includes T^l.

Moreover, observe that $S\subseteq\tilde{T}\backslash\Theta$ and $\Theta\subseteq T\subseteq\tilde{T}$, implying that $W(S)\leqslant V(\tilde{T}\backslash\Theta)=(\lambda(\tilde{T})-\lambda(\Theta))/(\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta)))$, where the inequality follows from the fact that the cost of the optimal coalition for S is a lower bound on the cost of any other coalition that includes S. To establish (3), we need to prove that $W(T^l)+W(S)\leqslant W(S^l)+W(T)$. Towards this end, note that

$$\begin{aligned}&W(T^l)+W(S)\leqslant V(\tilde{S}^l\cup\Theta)+V(\tilde{T}\backslash\Theta)\\&\quad=\frac{\lambda(\tilde{S}^l)+\lambda(\Theta)}{\mu(\tilde{S}^l)+\mu(\Theta)-(\lambda(\tilde{S}^l)+\lambda(\Theta))}\\&\qquad+\frac{\lambda(\tilde{T})-\lambda(\Theta)}{\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta))}.\end{aligned}$$

Therefore, it is sufficient to show instead that

$$\begin{aligned}&\frac{\lambda(\tilde{S}^l)}{\mu(\tilde{S}^l)-\lambda(\tilde{S}^l)}+\frac{\lambda(\tilde{T})}{\mu(\tilde{T})-\lambda(\tilde{T})}\\&\quad-\bigg(\frac{\lambda(\tilde{S}^l)+\lambda(\Theta)}{\mu(\tilde{S}^l)+\mu(\Theta)-(\lambda(\tilde{S}^l)+\lambda(\Theta))}\\&\qquad+\frac{\lambda(\tilde{T})-\lambda(\Theta)}{\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta))}\bigg)\geqslant 0.\end{aligned}$$

Note that the sign of the left-hand side of the above is as the sign of

$$\begin{aligned}&\lambda(\tilde{S}^l)[\mu(\tilde{T})-\lambda(\tilde{T})][\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta))]\\&\cdot[\mu(\Theta)-\lambda(\Theta)]-\lambda(\tilde{T})[\mu(\tilde{S}^l)-\lambda(\tilde{S}^l)]\\&\cdot[\mu(\tilde{S}^l)+\mu(\Theta)-(\lambda(\tilde{S}^l)+\lambda(\Theta))][\mu(\Theta)-\lambda(\Theta)]\\&+\lambda(\Theta)[\mu(\tilde{S}^l)-\lambda(\tilde{S}^l)][\mu(\tilde{T})-\lambda(\tilde{T})]\\&\cdot[\mu(\tilde{S}^l)-\mu(\tilde{T})+2\mu(\Theta)-(\lambda(\tilde{S}^l)-\lambda(\tilde{T})+2\lambda(\Theta))]\end{aligned}$$

Divide this expression by the positive constant

$$[\mu(\Theta)-\lambda(\Theta)][\mu(\tilde{T})-\lambda(\tilde{T})][\mu(\tilde{S}^l)-\lambda(\tilde{S}^l)]$$

and it remains to show that the following expression is nonnegative:

$$\begin{aligned}&W(S^l)[\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta))]\\&-W(T)[\mu(\tilde{S}^l)+\mu(\Theta)-(\lambda(\tilde{S}^l)+\lambda(\Theta))]\\&+V(\Theta)[\mu(\tilde{S}^l)-\mu(\tilde{T})+2\mu(\Theta)-(\lambda(\tilde{S}^l)-\lambda(\tilde{T})+2\lambda(\Theta))],\end{aligned}$$

which equals

$$\begin{aligned}&-[V(\Theta)-W(S^l)][\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta))]\\&+[V(\Theta)-W(T)][\mu(\tilde{S}^l)+\mu(\Theta)-(\lambda(\tilde{S}^l)+\lambda(\Theta))].\end{aligned}$$

Utilizing the equality $V(\Theta)-W(S^l)=V(\Theta)-W(T)+W(T)-W(S^l)$, we need to show that

$$\begin{aligned}&[V(\Theta)-W(T)][\mu(\tilde{S}^l)-\mu(\tilde{T})+2\mu(\Theta)\\&\qquad\qquad-(\lambda(\tilde{S}^l)-\lambda(\tilde{T})+2\lambda(\Theta))]\\&+[W(S^l)-W(T)][\mu(\tilde{T})-\mu(\Theta)-(\lambda(\tilde{T})-\lambda(\Theta))]\geqslant 0.\end{aligned}$$

Using the first part of Lemma 3 repetitively, we get that $W(S^l)=V(\tilde{S}^l)<V(\max\Theta)\leqslant V(\Theta)$, and that $W(S^l)=V(\tilde{S}^l)\geqslant V(\{b_{S^l}\})\geqslant V(\{c_T\})>V(\tilde{T})=W(T)$. This implies that $V(\Theta)>W(S^l)>W(T)$. This fact, coupled with the fact that $\Theta\subseteq T$, $\tilde{T}\subseteq\tilde{S}^l\cup\Theta$ (which holds because in this case we assume that $b_{S^l}\leqslant c_T$), and $\mu(\Theta)>\lambda(\Theta)$, we get that the last left-hand side is nonnegative as required. □

Acknowledgments

This research was partially funded by the Israel Science Foundation grant 401/08. The research of the first author was also partially funded by the Israeli Institute for Business Research. The authors thank Xin Chen for pointing out to them that the allocation $x_i=\lambda_i/(\xi(N)-\lambda(N))$ is in the core of the game (N,V) for the case $\alpha=1$. Part of the research leading to this paper was done while the second author was at Operations Management and Econometrics, Faculty of Economics and Business, University of Sydney.

References

Anily, S., M. Haviv. 2007. Cost-allocation for the first order interaction joint replenishment model. *Oper. Res.* **55**(2) 292–302.

Anupindi, R., Y. Bassok, E. Zemel. 2001. A general framework for the study of the decentralized distribution systems. *Manufacturing Service Oper. Management* **3**(4) 349–368.

Bell, C. E., S. Stidham Jr. 1983. Individual versus social optimization in allocation of customers to alternative servers. *Management Sci.* **29**(7) 831–839.

Chen, X., J. Zhang. 2009. Duality approaches to economic lot sizing games. Submitted.

Chun, Y. 2006a. A pessimistic approach to the queueing problem. *Math. Social Sci.* **51** 171–181.

Chun, Y. 2006b. No-envy in queueing problems. *Econom. Theory* **29**(1) 151–162.

Dror, M., B. C. Hartman. 2005. Shipment consolidation: Who pays for it and how much? *Management Sci.* **53** 78–87.

Gerchak, Y., D. Gupta. 1991. Technical note: On apportioning costs to consumers in centralized continuous review inventory systems. *J. Oper. Management* **10**(4) 546–551.

Gonzalez, P., C. Herrero. 2004. Optimal sharing of surgical costs in the presence of queues. *Math. Methods Oper. Res.* **59** 435–446.

Granot, D., G. Sošić. 2003. A three-stage model for a decentralized distribution system of retailers. *Oper. Res.* **51**(5) 771–784.

Guardiola, L. A., A. Meca, J. Puerto. 2009. Production-inventory games: A new class of totally balanced combinatorial optimization games. *Games Econom. Behav.* **65**(1) 205–219. doi:10.1016/j.geb.2007.02.003.

Hartman, B., M. Dror. 1996. Cost allocation in continuous review inventory models. *Naval Res. Logist.* **43**(4) 549–561.

Hartman, B., M. Dror, M. Shaked. 2000. Cores of inventory centralization games. *Games Econom. Behav.* **31**(1) 26–49.

Haviv, M. 1999. The Aumann-Shapley price mechanism for allocating costs in congested systems. *Oper. Res. Lett.* **29**(5) 221–229.

Haviv, M., Y. Ritov. 1998. Externalities, tangible externalities and queueing disciplines. *Management Sci.* **44**(6) 850–856.

Isciishi, T. 1981. Super-modularity applications to convex games and to the greedy algorithm for LP. *J. Econom. Theory* **25**(2) 283–286.

Katta, A.-K., J. Sethuraman. 2006. Cooperation in queues. CORC technical reports, TR-2005-03, Columbia University, New York.

Maniquet, F. 2003. A characterization of the Shapley value in queueing problems. *J. Econom. Theory* **109**(1) 90–103.

Meca, A., J. Timmer, I. Garia-Jurado, P. Borm. 2004. Inventory games. *Eur. J. Oper. Res.* **156** 127–139.

Moulin, H. 1995. *Cooperative Microeconomics: A Game-Theoretic Introduction*. Princeton University Press, Princeton, NJ.

Muller, A., M. Scarsini, M. Shaked. 2002. The newsvendor game has a non-empty core. *Games Econom. Behav.* **38**(1) 118–126.

Osborne, M. J., A. Rubinstein. 1994. *A Course in Game Theory.* MIT Press, Cambridge, MA.

Ozen, U., J. Fransoo, H. W. Norde, M. Slikker. 2004. Cooperation between multiple newsvendors with warehouses. CentER discussion paper 2004-34, Tilburg University, Tilburg, The Netherlands.

Rudi, N., S. Kapur, D. Pyke. 2001. A two-location inventory model with transshipment and local decision making. *Management Sci.* **47**(12) 1668–1680.

Shapley, L. 1971. Cores of convex games. *Internat. J. Game Theory* **1**(1) 11–26.

Slikker, M., J. Fransoo, M. Wouters. 2005. Cooperation between multiple news-vendors with transshipments. *Eur. J. Oper. Res.* **167**(2) 370–380.

van den Heuvel, W., P. Borm, H. Hamers. 2007. Economic lot-sizing games. *Eur. J. Oper. Res.* **176**(2) 1117–1130.

Yu, Y., S. Benjaafar, Y. Gerchak. 2006. On service capacity pooling and cost sharing among independent firms. Working paper, Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN.

Zhang, J. 2009. Cost allocation for joint replenishment models. *Oper. Res.* **57** (1) 146–156.