POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 0, No. 0, xxxx–xxxx 2017, pp. 1–17
ISSN 1059-1478 | EISSN 1937-5956 | 17 | 00 | 0001
DOI 10.1111/poms.12708

Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games

Shoshana Anily
Coller School of Management, Tel Aviv Univerisity, Tel Aviv, Israel, anily@post.tau.ac.il

Moshe Haviv
Department of Statistics and the Federman Center for the Study of Rationality, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel, moshe.haviv@gmail.com

We consider production and service systems that consist of parallel lines of two types: (i) $M/M/1$ lines and (ii) lines that have no buffers (loss systems). Each line is assumed to be controlled by a dedicated supervisor. The management measures the effectiveness of the supervisors by the long run expected cost of their line. Unbalanced lines cause congestion and bottlenecks, large variation in output, unnecessary wastes and, ultimately, high operating costs. Thus, the supervisors are expected to join forces and reduce the cost of the whole system by applying line-balancing techniques, possibly combined with either strategic outsourcing or capacity reduction practices. By solving appropriate mathematical programming formulations, the policy that minimizes the long run expected cost of each of the parallel-lines system, is identified. The next question to be asked is how to allocate the new total cost of each system among the lines' supervisors so that the cooperation's stability is preserved. For that sake, we associate a cooperative game to each system and we investigate its core. We show that the cooperative games are reducible to market games and therefore they are totally balanced, that is, their core and the core of their subgames are non-empty. For each game a core cost allocation based on competitive equilibrium prices is identified.

Key words: cooperative games; core; queues; production/service; line balancing
History: Received: February 2016; Accepted: February 2017 by Michael Pinedo, after 2 revisions.

1. Introduction

Unbalanced lines in manufacturing systems and service systems cause congestion and bottlenecks that result in large variation in output, unnecessary wastes and, ultimately, high operating costs. Various line-balancing practices that help improving the efficiency of the system are known where the most common one is pooling certain inner resources and redistributing them optimally. A line-balancing policy, which is based solely on pooling inner resources is called a *domestic processing policy*. In other words, a line-balancing policy that uses all its capacity in-house, and all its demand is satisfied by activities that are performed in-house using this capacity, is a domestic processing policy. However, in the few last decades, contracting out some activities by manufacturers and service providers has also become widespread. Strategic outsourcing enables firms to maintain control of critical core production or service competencies by releasing some inner capacity and resources towards the performance of tasks in which the firm specializes in and where it has a competitive advantage over its rivals. Another practice that is often used by firms at times when the demand for their products declines is to reduce the capacity as maintenance of a too high capacity is expensive. By applying such practices, firms have achieved competitive edge and have substantially increased their productivity.

In the sequel, we describe the service/production systems considered here, and the motivation behind this research. The service systems consist of parallel servers where each server is responsible for serving a certain group of customers. The role of a server may be played by an individual person, a team of workers or by an automated mechanism. Examples for such services include a medical examination such as clinical breast exams for women, MRI screening, renewal of professional licences, etc. In the context of manufacturing, the systems consist of parallel production units. For the sake of generality, we refer to parallel *lines* rather than servers or machines. Each line is assumed to be a separate cost unit for accounting purposes and therefore it is assigned a dedicated supervisor who is responsible for the effectiveness of the line. At the end of each fiscal year, the management distributes bonuses to the supervisors based on their performance, where their performance is evaluated by the cost of their line so that the lower the cost is, the higher the bonus is. Suppose that the management

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

is interested in improving the system by using a certain combination of practices such as pooling and redistributing some resources, (e.g., demands or capacities), shutting off certain ineffective lines, outsourcing some demand or reducing some surplus capacity. Such practices may reduce the total cost, but they also affect the characteristics and the cost of the individual lines. For example, suppose that the lines are evaluated by their respective congestion cost and one of the supervisors prior the change improved her line by increasing its capacity and reducing its congestion, and now she is asked to transfer some of her surplus capacity to some other lines. By doing so, the total cost may go down while her direct congestion cost may go up. How will she be compensated for contributing her share to lower the total cost of the system while her own cost has increased? Without an appropriate compensation she might refuse to contribute her resources to the team. This is the type of questions that we ask in this study. Given a system, first we minimize its total cost by line-balancing techniques in an optimal way, and then we look for a fair scheme that allocates the total cost to the individual lines.

We consider systems that operate a number of nonidentical parallel lines, where each line is associated with its own exponential processing time, its own Poisson demand process and its own cost parameter. Units that arrive to a line whose buffer is full are discarded and lost forever. We consider systems that share the same size of buffers for all the lines, and in this research we focus on the two extreme cases where either all buffers are infinitely large or all buffers have a zero size. Processing of a unit by a line starts immediately if upon its arrival the line is idle. Otherwise, the units queue up in the line's buffer as long as the buffer is not full. Units are processed one-by-one by the lines according to a *First Come First Served* (FCFS) policy. Initially, the system on hand is assumed to use a domestic processing policy, that is, each line operates at its full capacity while servicing its demand. Thus we get two types of systems: The first consists of parallel $M/M/1$ lines where the cost of a line is its long run expected congestion cost. The second consists of parallel $M/M/1/1$ lines, where no buffers exist, and the cost of a line is its long run expected cost due to discarded units. In both cases, it is assumed that the cost of the whole system is additive in the costs of the individual lines.

$M/M/1$ lines are quite common in modeling both in service and manufacturing systems as, on one hand, they approximate numerous real models well and, on the other hand, there exists a wide body of knowledge that sheds light on their properties. In an $M/M/1$ line, all demand is eventually satisfied and the line is evaluated by its congestion cost. However, in practice, it is often the case that lines have finite buffers and therefore their long run expected cost should take into consideration both the congestion and the rate of discarded units. As a first step toward the study of parallel lines with general buffer size, we consider here systems of $M/M/1/1$ lines where no queues are accumulated and their cost is directly associated with the expected number of loss units.

The line-balancing techniques that we consider here are: (i) *unobservable routing* where units in the pooled arrival streams can be rerouted among all lines and (ii) *capacity sharing* where the total capacity is pooled and can be reassigned among the lines. Unobservable routing may be coupled with outsourcing, that is, some of the demand can be outsourced while the rest is routed optimally among the lines. The total cost of such a system is the cost of the balanced lines (congestion cost in parallel $M/M/1$ systems and cost of discarded units in parallel loss systems), plus the outsourcing cost. Capacity sharing may be coupled with reduction of capacity, so that the capacity that is left for in-house activities is optimally reassigned among the lines. In such a case, the total cost is the cost of the balanced lines, as described above, minus the savings due to capacity reduction.

The problem of minimizing the cost of a system by line-balancing, outsourcing and capacity reduction, can be formulated as a mathematical programming problem. The solution of such a problem generates both the optimal policy and the optimal cost. However, in many cases this is just the first step in a successful implementation of the optimal policy as the management might be interested in allocation of the total cost of the system among the lines as, for example, for the sake of applying a bonus scheme to the lines' dedicated supervisors. In order to achieve full cooperation while implementing the optimal line-balancing policy, the management needs to specify an allocation scheme of the optimal cost among the lines so that it will be accepted by all the supervisors. This is exactly the subject of the theory of *cooperative games with transferable utilities*. In section 2, some of the main concepts of this theory are presented, where here we briefly explain the general approach.

A cooperative game is defined by a given set of *players* and a *characteristic function*. The characteristic function is a set function that returns the cost of each *coalition*, e.g., subset of players. In our context, we regard the lines (or their dedicated supervisors) as the players, and the characteristic function returns the expected long run average cost for any coalition of players. Initially, the cost of the system is the sum of the costs of the individual players. If the set of players is partitioned into disjoint coalitions, then the cost of the system is the sum of the costs of all coalitions in the partition. Under certain conditions on the

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

characteristic function, players may be better off cooperating. In all the games that we consider, the form of the characteristic function ensures full cooperation among the players, so that any bargaining process among the players would probably end up in full cooperation, and in the formation of the *grand coalition*, that is, the coalition that consists of all players. Once that the grand coalition is formed, the next natural question is how to fairly allocate the total cost among the players in order to ensure the stability of the grand coalition in view of the fact that some players have a greater contribution to the grand coalition than others. Several concepts of fairness have been proposed in the literature. The most appealing one that we adopt here, is the *core*: A cost allocation vector is in the core of the game if it is *efficient*, that is, the sum of its entries equals the cost of the grand coalition, and if, in addition, for any coalition of players, the total cost allocated to its players is bounded from above by the cost of the coalition if its players would join forces and abandon the grand coalition. That means, that a cost vector is in the core if and only if no coalition has an incentive to defect and play by itself. We note here that a cost allocation vector is not necessarily non-negative, that is, there may exist players that are allocated a negative cost. This may occur in games where "valuable" players exist (in our context, lines that have a large capacity and a low arrival rate) and the other players may agree to pay the "valuable" players in order to convince them to cooperate, as they may help reducing substantially the total cost. In general, the core is either empty, or infinitely large or it consists of a single cost allocation. Though, the definition of the core sounds reasonable, characterizing the core may be an intricate task. Indeed, this issue coupled with the possibility that the core is empty, makes the problem of finding a core allocation a real challenge in some games, let alone characterizing the whole core.

In this study, we consider the problems of line-balancing by unobservable routing jointly with outsourcing, and of line-balancing by capacity sharing and capacity reduction, in parallel $M/M/1$ and $M/M/1/1$ systems. For each problem the optimal policy and the long run average cost are identified. The problems are then formulated as cooperative games and for each game a core cost allocation is identified.

The rest of the study is organized as follows: In section 2 we state the main definitions and prerequisites on cooperative games, and we present the class of *market games*. In section 3, we consider parallel $M/M/1$ lines: we find the optimal domestic processing policy for each of the two versions of the problem and we define the respective line-balancing games for an extended version of the problem where either demand can be outsourced or capacity can be reduced. In section 4, we consider the same questions on parallel $M/M/1/1$ lines. All the four games are shown to be reducible to market games, proving that they are totally balanced. A core allocation for each game based on competitive equilibrium pricing is found. Section 5 concludes the study.

2. Review on Cooperative Games

A general *cooperative game with transferable utility* is defined by a pair $G = (N, c)$, where the set $N = \{1, 2, \ldots, n\}$ consists of n players, any subset S of N, $\varnothing \subseteq S \subseteq N$, is called a *coalition*, where N itself is called the *grand coalition* and each coalition S is associated with a real value denoted by $c(S)$, where $c(\varnothing) = 0$. The value $c(S)$ is the total cost inflicted on the members of coalition S if its members, and only its members, cooperate. The set function $c : 2^N \rightarrow \Re$ is called a *characteristic function*. The total cost incurred by all players of N that partition into m disjoint coalitions, that is, $S_1 \cup S_2 \cup \ldots \cup S_m = N$, $1 \leq m \leq n$, is $\sum_{i=1}^{m} c(S_i)$. Note that in general cooperative games with transferable utility, the total cost is not necessarily additive in the coalitions but the additive form is the most conventional form that used in the literature. In the line-balancing games considered here, the individual lines play the role of the players and the characteristic function maps each coalition to its long run expected total cost, as obtained by applying the line-balancing procedure on its members.

Next we review the main concepts in cooperative games that are relevant to this study. Given a game, the first question is whether the grand coalition is the socially optimal formation of coalitions, that is, whether $c(N) \leq \sum_{i=1}^{m} c(S_i)$ for any partition to disjoint coalitions $S_1 \cup S_2 \cup \ldots \cup S_m = N, 1 \leq m \leq n$. A sufficient condition for full cooperation is subadditivity of its characteristic function: A game $G = (N, c)$ is called *subadditive* if for any two coalitions S and T, $c(S \cup T) \leq c(S) + c(T)$. Subadditive games bear the concept of *economies of scope*, that is, when each player, or set of players, contributes its own skills and resources, the total cost is no greater than the sum of the costs of the individual parts. On top of forming the grand coalition, it is necessary to establish a way that allocates the cost $c(N)$ among the players, so that no group of players may resist the cooperation and decide to act alone. Several concepts of stability have been proposed in the literature. The most appealing is the *core*: A vector $x \in \Re^n$ is said to be *efficient* if $\sum_{i=1}^{n} x_i = c(N)$, and it is said to be a *core cost allocation* of the game if it is efficient and if it satisfies the 2^{n-2} *stand-alone* inequalities, namely, $\sum_{i \in S} x_i \leq c(S)$ for any $S \subset N$.

The collection of all core allocations, called the *core* of the game, forms a polyhedron in $\Re^n$ as it is defined

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

by a set of linear constraints with n decision variables. As the number of constraints that define the core is exponential in n, finding a core allocation for a given game may be, in general, an intricate task. Indeed, this issue coupled with the possibility that the core is empty, makes the problem of finding a core allocation a real challenge in some games. Moreover, even if one can prove the non-emptiness of the core, the question of finding a cost allocation in the core may be non-trivial, let alone characterizing the whole core.

A cooperative game $G = (N, c)$ is said to be *balanced* if its core is non-empty, and *totally balanced* if its core and the cores of all its subgames are non-empty. Subadditivity is a necessary condition for total balancedness as if there existed disjoint coalitions S and T for which $c(S) + c(T) < c(S \cup T)$, the subgame $(S \cup T, V)$ would have an empty core since any efficient allocation of $c(S \cup T)$ among the players of $S \cup T$ will be objected by at least one of the coalitions, S or T.

Some papers have considered resource pooling in the context of cooperative games, see e.g., Anily and Haviv (2010), Chakravarthy (2016), Karsten (2013), Karsten et al. (2009, 2011), Timmer and Scheinhardt (2013) and Yu et al. (2015). Directly related to this study is Timmer and Scheinhardt (2013), which proves that the domestic processing capacity sharing game with identical cost parameters across all lines, is totally balanced and it identifies a specific cost allocation in the core. We generalize the game by allowing line-dependent cost parameters in addition to capacity reduction.

The literature provides two main general conditions that are sufficient in order establish the total balancedness of a game.

CONDITION 1. *A game $G = (N, c)$ is a concave game if its characteristic function is concave, meaning that for any two coalitions $S, T \subseteq N$, $c(S \cup T) + c(S \cap T) \leq c(S) + c(T)$. Concave games are, clearly, subadditive but not the other way around. It was shown in Shapley* (1971) *that the core of a concave game possesses n! extreme points, each of which being the vector of marginal contribution of the players for a different permutation of the players.*

CONDITION 2. *A market game, see e.g., Chapter* 13 *in Osborne and Rubinstein* (1994)*, is defined as follows: Suppose there are ℓ types of inputs. An input vector is a non-negative vector in $\Re_+^\ell$. Each of the n players possesses an initial commitment vector $w_i \in \Re_+^\ell$, $1 \leq i \leq n$, which states a nonnegative quantity for each input. Moreover, each player is associated with a continuous and convex cost function $g_i : \Re_+^\ell \to \Re_+$, $1 \leq i \leq n$. A profile $(z_i)_{i \in N}$ of input vectors for which $\sum_{i \in N} z_i = \sum_{i \in N} w_i$ is an allocation. The game is such that a coalition S of players looks for an optimal way to redistribute its members' total commitments among its members in order to get a profile $(z_i)_{i \in S}$ of input vectors so as the sum of the costs across the members of S is minimized. Formally, for any $\varnothing \subseteq S \subseteq N$,*

$$c(S) = \min \Big\{ \sum_{i \in S} g_i(z_i) : z_i \in \Re_+^\ell, \; i \in S \text{ and } \sum_{i \in S} z_i = \sum_{i \in S} w_i \Big\}. \tag{1}$$

Market games are not necessarily concave, but they are well-known to be totally balanced, see Peleg and Sudholter (2007), Corollary 3.2.4. Unlike concave games whose core is fully characterized and has a closed form (see Condition 1), just a single core allocation $(x_1, \ldots, x_n)$, given in equation (2) below, which is based on competitive equilibrium prices, is known for a general market game, (see Osborne and Rubinstein (1994, p. 266):

$$x_i = g_i(z_i^*) - \Theta(z_i^* - z_i) \quad \text{for } i \in N, \tag{2}$$

where Θ is the Lagrange multiplier of the constraint in equation (1), and $(z_i^*)_{i=1}^n$ signifies the optimal input to each player in equation (1).

In fact, Shapley and Shubik (1969) proves that a game is a market game if and only if it is totally balanced. In particular, any concave game is a market game. However, if a game is not naturally formulated as a market game (see equation (1)), then the task of reformulating it as a market game (or showing that such a formulation does not exist), may be as intricate as proving directly that it is totally balanced (or that it is not). Thus, it seems that except for games that are either originally stated as market games, or are easily transformed to market games, this approach has its limits. We show that each of the line-balancing games described here, is transformable to a market game, enabling us to derive the cost allocation that is based on competitive equilibrium prices.

Anily and Haviv (2010) analyzes the most basic $M/M/1$ service pooling game, where cooperation among $M/M/1$ lines generates a new $M/M/1$ line whose arrival stream is the union of the individual streams, its capacity is the sum of the individual capacities and the characteristic function returns the long run expected congestion. The game is proved to be totally balanced though it is neither concave nor monotone, and the nonnegative part of its core is fully characterized. Anily and Haviv (2014) defines a large class of games, called *regular games*, that contains most cooperative games in service and operations management including the above mentioned $M/M/1$ service pooling game and the line-balancing games considered in this study. Anily (2017) proposes a sufficient condition for total balancedness of a sub-class of

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

subadditive homogenous of degree 0 regular games that is applicable to the $M/M/1$ service pooling game. Anily and Haviv (2014) proves that a subadditive and homogenous of degree 1 regular game is totally balanced, a result that is applicable to all the line-balancing games considered here. Unlike Conditions 1 and 2, that not only provide sufficient conditions for the non-emptiness of the core but they also indicate a methodology to compute a core allocation, no core allocation is known for games that satisfy the two sufficient conditions regarding regular homogenous of degree 1 (0) described above.

3. Line Balancing Games of Parallel $M/M/1$ Line Systems

In this section we discuss line balancing models that consist of n non-identical parallel lines, $N = \{1, \ldots, n\}$, with infinite buffer where each line is associated with its own Poisson arrival process of demands, and its service time, which is exponentially distributed. Line $i \in N$ is associated with a mean service rate $\mu_i > 0$, and a mean arrival rate $\lambda_i \geq 0$, where $\lambda_i < \mu_i$. In such a system the congestion cost is of concern. The cost per unit of congestion per unit of time on line $i \in N$ is $\alpha_i > 0$, implying that its long run average congestion cost is $\frac{\alpha_i \lambda_i}{\mu_i - \lambda_i}$. The total cost is additive in the cost of the lines. In the following, we use the notation $\lambda(S) = \sum_{i \in S} \lambda_i$ and $\mu(S) = \sum_{i \in S} \mu_i$ for any subset $\varnothing \subseteq S \subseteq N$.

Within the class of domestic processing policies, that is, where the total capacity $\mu(N)$ is used to serve the total demand rate $\lambda(N)$ in-house without using outsourcing or capacity reduction, two possible improvement schemes of parallel $M/M/1$ line systems have been proposed in the literature: (i) the capacities of the individual lines are preserved at their original levels, but the pull of the arrival streams of rate $\lambda(N)$, can be rerouted among the lines. Such a situation may occur in a production system where the lines are identical machines whose production rate is fixed and machine dependent, but the lines' input rates are decision variables. This version of the problem is called the *unobservable routing problem*. (ii) the individual streams of arrivals are kept as given, but the total pooled capacity $\mu(N)$ can be reassigned among the lines. Such a case may occur, for example, if the last production stage of identical parallel machines is painting components in different colors, where one machine paints in blue, another one in red, etc. The input rate to each machine is the given demand rate for components of a certain color. The capacities of the machines can be adjusted by the corresponding demand rates. This version of the problem, given that the total capacity $\mu(N)$ is reassigned among the machines, is called the *capacity sharing* problem. We consider a more general class of line balancing policies that contains the class of domestic processing policies, where unobservable routing can be coupled with outsourcing some demand, and capacity sharing can be coupled with capacity reduction.

In each of the next two subsections, we consider the parallel $M/M/1$ line balancing optimal policy and the respective cooperative game under unobservable routing and under capacity sharing. Each subsection introduces the corresponding optimal domestic processing policy. Then we extend the class of policies to allow either outsourcing some demand at a linear cost, if demands are pooled, or reducing some capacity for linear savings, if capacities are pooled. For each model we formulate and solve a mathematical programming problem that balances the lines optimally by generating for the unobservable routing problems the demand to be outsourced and the demand directed to each line, and for the capacity sharing problems, the excess capacity that is reduced as well as the capacity that is assigned to each line. The optimal solutions demonstrate the dependence of the optimal cost in the various parameters of the systems. Then, we consider the question of how to fairly allocate the optimal cost among the supervisors of the lines, given that each line is assigned its own dedicated supervisor, so that no individual supervisor or set of supervisors has an incentive to defect from the full cooperation. Note that usually the supervisors that have an incentive to break away from full cooperation, are the ones that seem to subsidize the others. Usually, those are the most efficient supervisors that the outcome of cooperation makes them regarded as less efficient as they are asked to bear a greater portion of the total load of the system in order to optimize the efficiency of the whole system.

3.1. The Unobservable Routing with Outsourcing Game

The optimal domestic processing policy for the unobservable routing problem for parallel $M/M/1$ lines for the case where the cost per unit of congestion is the same for all lines, has been derived in Bell and Stidham (1983) (see also Hassin and Haviv (2003, p. 65), while the case of line-dependent congestion cost parameters, that we present next, has been solved in Altman et al. (2011). Thereafter, we derive the optimal policy for the line-dependent congestion cost parameters with the option of outsourcing demand.

Assume that the lines are indexed in a non-decreasing order of $\frac{\alpha_i}{\mu_i}$, that is, $\alpha_1/\mu_1 \leq \alpha_2/\mu_2 \ldots \leq \alpha_n/\mu_n$. As we are going to see, the structure of the optimal solution is such that a consecutive set of the highest indexed lines, might be idle, that is, no demand will be directed to these lines.

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

Let denote the arrival rate to be assigned to line i by z_i,$0 \le z_i < \mu_i$. Let $\tau_i(z)$ represent the congestion cost of line i for arrival rate $z < \mu_i$, where

$$\tau_i(z) = \frac{\alpha_i z}{\mu_i - z}. \tag{3}$$

The corresponding optimal domestic processing policy problem is defined by:

$$c(N) = \min\Big\{ \sum_{i=1}^{n} \tau_i(z) : \sum_{i=1}^{n} z_i = \lambda(N) \quad \text{and } z_i \ge 0 \text{ for } i = 1, \ldots, n \Big\}. \tag{4}$$

The function $\tau_i(z)$ is an increasing and convex function of z. The marginal cost of directing an infinitesimal demand to line i is $\frac{d\tau_i(z)}{dz}|_{z=0^+} = \frac{\alpha_i}{\mu_i}$. Thus, the optimal policy can be obtained by gradually increasing, starting from zero, the demand directed to the first line until the marginal congestion cost on this line reaches the marginal congestion cost of directing an infinitesimal demand to the second line. Subsequently, the demands on the two first lines are gradually increased until the marginal costs on these two lines reach the marginal cost of directing an infinitesimal demand to the third line. The assignment process continues until all the total demand $\lambda(N)$ is assigned to lines. The assignment process may end while some of the highest indexed lines are not used. Denote the index of the last open line by i^*. Let z_i^* be the optimal arrival rate directed to line i, where $z_i^* = 0$ for $i > i^*$. The Lagrange multiplier of the equality constraint in equation (4) is denoted by Ψ. Thus, for $i \le i^*$, $\frac{d\tau_i(z)}{dz}|_{z=z_i^*} = \Psi$, and for $i > i^*$, $\frac{d\tau_i(z)}{dz}|_{z=z_i^*=0} = \frac{\alpha_i}{\mu_i} \ge \Psi$, where

$$i^* = \min\left\{ i \in N : \frac{\alpha_{i+1}}{\mu_{i+1}} \ge \frac{\left(\sum_{j=1}^{i} \sqrt{\alpha_j \mu_j}\right)^2}{\left(\sum_{j=1}^{i} \mu_j - \lambda(N)\right)^2} \right\}, \tag{5}$$

$$\Psi = \frac{\left(\sum_{j=1}^{i^*} \sqrt{\alpha_j \mu_j}\right)^2}{\left(\sum_{j=1}^{i^*} \mu_j - \lambda(N)\right)^2}, \tag{6}$$

and the optimal congestion level is shown by minimal algebra to equal

$$c(N) = \frac{\left(\sum_{i=1}^{i^*} \sqrt{\alpha_i \mu_i}\right)^2}{\sum_{i=1}^{i^*} \mu_i - \lambda(N)} - \sum_{k=1}^{i^*} \alpha_i. \tag{7}$$

The optimal routing rate to any open line is

$$z_i^* = \mu_i - \Big(\sum_{j=1}^{i^*} \mu_j - \lambda(N) \Big) \frac{\sqrt{\alpha_i \mu_i}}{\sum_{j=1}^{i^*} \sqrt{\alpha_j \mu_j}}, \quad 1 \le i \le i^*. \tag{8}$$

Next, we generalize the above problem, by allowing to outsource some or all of the demand while applying a line balancing policy. The unit outsourcing cost rate is set to 1, and the line-dependent congestion cost parameters are scaled accordingly. Define a game $G = (N, \tilde{c})$ where each coalition $\varnothing \subseteq S \subseteq N$ is associated with a cost $\tilde{c}(S)$ that represents the optimal expected long run congestion and outsourcing cost incurred by a demand rate of $\lambda(S)$ that is met by a policy that combines outsourcing and processing by the lines of S where line $i \in S$ is associated with a capacity μ_i.

In order to prove that the unobservable routing with outsourcing game $G = (N, \tilde{c})$ in parallel $M/M/1$ lines is totally balanced, we reduce it to a market game. For that sake, let the function $\phi_i(\lambda)$, for $\lambda \ge 0$, given by equation (9), represent the optimal expected long run congestion and outsourcing cost of line $i \in N$ that faces a demand a rate of λ.

$$\phi_i(\lambda) = \min\{\tau_i(z) + \lambda - z| \ \ 0 \le z \le \lambda\}. \tag{9}$$

Line i, $i \in N$, is better off processing units than outsourcing as long as its marginal cost is smaller than 1. Thus, let $\bar{z}_i$ for lines $i \in N$, whose marginal cost at $z_i = 0$ is smaller than 1, to be the rate at which the marginal cost is 1. Otherwise, let $\bar{z}_i = 0$.

Proposition 1. *The optimal policy that minimizes the sum of congestion and outsourcing costs of an M/M/1 line i, $i \in N$, with capacity $\mu_i > 0$, and congestion cost rate α_i, is unique: The line processes a demand rate of at most $\bar{z}_i$, and the rest, if positive, is outsourced, where*

$$\bar{z}_i = \max\{0, \mu_i - \sqrt{\alpha_i \mu_i}\}. \tag{10}$$

Proof. The solution to the equation $\frac{d\tau_i(z)}{dz} = \frac{\alpha_i \mu_i}{(\mu_i - z_i)^2} = 1$ is $\mu_i - \sqrt{\alpha_i \mu_i}$. If this value is positive then $\bar{z}_i = \mu_i - \sqrt{\alpha_i \mu_i}$. Otherwise, it is zero. The uniqueness of the policy is due to the fact that the function $\tau_i(z)$, see equation (3), is strictly convex in $(0, \bar{z}_i)$. □

Proposition 1 implies that any line $i \in N$ with $\frac{\alpha_i}{\mu_i} \ge 1$, is closed at optimality, and its demand is either processed by other, cheaper, lines of N or it is outsourced. In particular, if line i with $\frac{\alpha_i}{\mu_i} \ge 1$, is a single line system, $\phi_i(\lambda_i) = \lambda_i$. Otherwise, namely if $\frac{\alpha_i}{\mu_i} < 1$, line i processes a number of units that does not exceed $\bar{z}_i$ and if line i is a single line system, the remaining demand, if positive, is outsourced. To summarize, $\phi_i(\lambda_i) = \tau_i(\min\{\lambda_i, \bar{z}_i\}) + \max\{\lambda_i - \bar{z}_i, 0\}$. However, when optimizing the total cost of a multi line system, it is possible that some lines with $\frac{\alpha_i}{\mu_i} < 1$, are also closed as it might be possible to process the total demand rate $\lambda(N)$ on cheaper lines.

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

The optimal unobservable routing with ousourcing cost in parallel $M/M/1$ lines of any coalition $\varnothing \subseteq S \subseteq N$, $\tilde{c}(S)$, is defined by

$$\tilde{c}(S) = \min\Big\{ \sum_{i\in S} \phi_i(z_i) : \sum_{i\in S} z_i = \lambda(S) \quad \text{and } z_i \geq 0 \text{ for } i \in S \Big\}. \tag{11}$$

The cost of the grand coalition, $\tilde{c}(N)$, is obtained by substituting S by N in equation (11).

THEOREM 1. *The unobservable routing with outsourcing game $G = (N, \tilde{c})$, on parallel $M/M/1$ lines system, where the characteristic function $\tilde{c}$ is defined in equation* (11)*, is a market game.*

PROOF. The characteristic function $\tilde{c} : 2^N \to \Re$ obeys the requirements of a market game, see equation (1), as the functions $\phi_i(z), z \geq 0,\ 1 \leq i \leq n$, are convex. □

In view of Theorem 1 and Condition 2, the unobservable routing with outsourcing game $G = (N, \tilde{c})$, on parallel $M/M/1$ lines system, is totally balanced and the cost allocation based on competitive equilibrium prices is in its core.

In the following, we investigate structural properties of the optimal solution of the grand coalition cost $\tilde{c}(N)$, see equation (11) for $S = N$, and we derive the core cost allocation based on competitive equilibrium prices for the game $G = (N, \tilde{c})$. Let Θ be the Lagrange multiplier of the equality constraint, and z_i^*, $i \in N$, be the optimal solution of the optimization problem (11) for $S = N$. Clearly, $\Theta \leq 1$ as the marginal cost of increasing $\lambda(N)$ is bounded by the outsorcing cost rate that equals 1. In fact,

$$\Theta = \min\{\Psi, 1\}, \tag{12}$$

where Ψ is the Lagrange multiplier of the equality constraint in equation (4), that is, the Lagrange multiplier for the routing problem without the outsourcing option. Recall that the lines are indexed in a non-decreasing order of the ratio between the cost rate and the capacity. The set of open lines for problem $\tilde{c}(N)$, see equation (11) by substituting S by N, is of the form $\{1, \ldots, i^o\}$, where

$$i^o = \min\left\{ i \in N : \frac{\alpha_{i+1}}{\mu_{i+1}} \geq \min\left\{ 1, \frac{\left(\sum_{j=1}^{i} \sqrt{\alpha_j \mu_j}\right)^2}{\left(\sum_{j=1}^{i} \mu_j - \lambda(N)\right)^2} \right\} \right\}. \tag{13}$$

By definition, $i^o \leq i^*$ where i^* is defined in equation (5). If outsourcing is not used by N, then $i^o = i^*$. The following lemma proves that the optimal number of units routed to each line is unique.

LEMMA 1. *An optimal solution $(z_1^*, \ldots, z_n^*)$ to problem* (11) *for $S = N$, satisfies one of the following two cases:*

- *for all $i \in N$, $z_i^* < \bar{z}_i$ or $z_i^* = 0$. Or,*
- *for all $i \in N$, $z_i^* \geq \bar{z}_i$.*

In addition, there exists a unique optimal routing of units to the lines, where the routing rate to line $i \in N$ is $\min\{z_i^*, \bar{z}_i\}$, and the outsourcing rate is $\lambda(N) - \sum_{i\in N} \min\{z_i^*, \bar{z}_i\}$.

PROOF. The convexity of the functions $\phi_i, i \in N$, in equation (11), implies that at optimality $\tilde{c}(N)$ satisfies the following properties: (i) the marginal cost of all lines with $z_i^* > 0$ is the same; (ii) the marginal cost of all lines with $z_i^* = 0$ is at least as high as the cost of the former group of lines; and (iii) the marginal cost of all lines is bounded from above by 1, which is the outsourcing cost rate. The Lagrange multiplier Θ of the equality constraint in equation (11) where $S = N$ is, in fact, the marginal cost of lines having $z_i^* > 0$. If there exists $j \in N$ such that $0 < z_j^* < \bar{z}_j$, then $\Theta < 1$, and the marginal congestion cost of all lines with $z_i^* > 0$, is Θ, where the marginal cost of lines with $z_i^* = 0$, is at least Θ. In such a case ousourcing is not used by N, and for each line $i = 1, \ldots, n$, z_i^* is the optimal rate of units processed by line i, that is, $z_i^* < \bar{z}_i$ or $z_i^* = 0$. In this case, the strict convexity of the functions $\phi_i(\lambda)$ in the range $\lambda \in [0, \bar{z}_i)$ implies a unique vector $(z_1^*, \ldots, z_n^*)$, which coincides with equation (8).

On the other hand, if there exists a line $j \in N$ such that $z_j^* \geq \bar{z}_j$, then $\Theta = 1$, implying that $z_j^* - \bar{z}_j$ units out of z_j^* are outsourced. Therefore, the marginal cost of increasing z_i^* for all lines $i \in N$, is also equal to $\Theta = 1$, meaning that $z_i^* \geq \bar{z}_i$ for all $i \in N$. At optimality, this solution means that each open line $i \leq i^o$ receives a rate of exactly $\bar{z}_i$ units of demand, and the total remaining rate of $\lambda(N) - \sum_{i=1}^{n} \bar{z}_i$, is outsourced. Thus, in a case the option of outsourcing is actually realized, the vector $(z_1^*, \ldots, z_n^*)$ that solves $\tilde{c}(N)$ is not unique but the optimal routing of units to the lines is unique. □

The next theorem specifies explicitly the core cost allocation $(x_i)_{i=1}^n$ based on competitive equilibrium prices, see equation (2). One of the interesting properties characterizing market games is that the form of the competitive equilibrium prices core allocation depends only on $\tilde{c}(N)$, rather than $\tilde{c}(S)$ for all coalitions $\varnothing \subseteq S \subseteq N$. In the context of the unobservable routing and outsourcing games, the vector $(x_i)_{i=1}^n$ depends only on whether or not the grand coalition outsources, and not on whether or not any of the other $2^n - 2$ sub-coalitions outsources. Recall that Θ is the Lagrange multiplier of the equality constraint in equation (11) for $S = N$.

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

THEOREM 2. *The optimal solution of the grand coalition and the competitive equilibrium prices core allocation* $(x_i)_{i=1}^n$ *of the unobservable routing with outsourcing game* $(N, \tilde{c})$ *are:*

- *If* $\Theta < 1$ *or if* $\Theta = 1$ *and* $\lambda(N) - \sum_{i=1}^{i^o} \bar{z}_i = 0$, *then outsourcing is not used by N. The last open line is* $i^o = i^*$, *where* i^* *is defined in equation* (5), *the optimal routing* z_i^* *to line* $i \in N$ *is given in equation* (8), *and the optimal cost* $\tilde{c}(N)$ *is given in equation* (7). *Finally, for* $i \le i^*$:

$$x_i = -(\mu_i - \lambda_i)\frac{\left(\sum_{k=1}^{i^*}\sqrt{\alpha_k\mu_k}\right)^2}{\left(\sum_{k=1}^{i^*}\mu_k - \lambda(N)\right)^2} + 2\sqrt{\alpha_i\mu_i}\frac{\sum_{k=1}^{i^*}\sqrt{\alpha_k\mu_k}}{\sum_{k=1}^{i^*}\mu_k - \lambda(N)} - \alpha_i\ , \quad 1 \le i \le i^*$$

and

$$x_i = \lambda_i \frac{\left(\sum_{k=1}^{i^*}\sqrt{\alpha_k\mu_k}\right)^2}{\left(\sum_{k=1}^{i^*}\mu_k - \lambda(N)\right)^2}, \quad i^* + 1 \le i \le n.$$

- *Otherwise, outsourcing is utilized by N. The last open line* i^o *is defined in equation* (13), *the optimal routing to lines* $i \in N$ *is* $z_i^* = \bar{z}_i$, *see Proposition* 1, *and the optimal cost is* $\tilde{c}(N) = \lambda(N) - \sum_{i \le i^o}\left(\sqrt{\mu_i} - \sqrt{\alpha_i}\right)^2$. *Finally, the competitive equilibrium prices core allocation is*

$$x_i = \lambda_i - \left(\sqrt{\mu_i} - \sqrt{\alpha_i}\right)^2, \quad 1 \le i \le i^o,$$

and

$$x_i = \lambda_i, \quad i^o + 1 \le i \le n.$$

PROOF. The form of the competitive equilibrium prices core allocation is given in equation (2), where Θ is the Lagrange multiplier of the equality constraint in equation (11) for $S = N$, and $(z_i^*)_{i=1}^n$ is a corresponding optimal routing to the problem. The proof of the two cases is as follows:

- If $\Theta < 1$, or if $\Theta = 1$ and $\lambda(N) - \sum_{i=1}^{i^o} \bar{z}_i = 0$, then the optimal policy is a domestic processing policy analyzed in Altman et al. (2011) and described at the beginning of this subsection. As the outsourcing option is not used, $\Theta = \Psi$, where Ψ is the Lagrange multiplier of the equality constraint of the optimization problem (4) whose value is defined in equation (6). In particular, line i^*, defined in equation (5), is the last open line, and the unique optimal routing rate z_i^*, to any line $i \le i^*$, is given in equation (8). In order to complete the proof of this item, we calculate the competitive equilibrium prices core cost allocation $(x_1, \ldots, x_n)$, see equation (2), by substituting the values of Θ and z_i^* for $i \in N$, into $x_i = \phi_i(z_i^*) - \Theta(z_i^* - \lambda_i) = \frac{\alpha_i z_i^*}{\mu_i - z_i^*} - \Theta(z_i^* - \lambda_i)$.
- Otherwise, outsourcing is used, implying that $\Theta = 1$, the last open line is i^o given in equation (13), the unique optimal routing rate to any line $i \le i^o$ is given by $\bar{z}_i$, see Proposition 1, and any optimal solution $(z_1^*, \ldots, z_n^*)$ to the optimization problem (11) for $S = N$, satisfies $z_i^* > \bar{z}_i$ for $i \le i^o$, $z_i^* = 0$ for $i = i^o + 1, \ldots, n$, and $\sum_{i \in N} z_i^* = \lambda(N)$. In particular, in any such solution, the rate at which units are outsourced is $\lambda(N) - \sum_{i=1}^{i^o} \bar{z}_i > 0$. Thus, any optimal solution of problem (11) for $S = N$, is of the following form: $z_i^* = \bar{z}_i + \delta_i$, where $\delta_i > 0$ for $i = 1, \ldots, i^o$, $z_i^* = 0$ for $i = i^o + 1, \ldots, n$, and $\sum_{i=1}^{i^o} \delta_i = \lambda(N) - \sum_{i=1}^{i^o} \bar{z}_i$. Substituting into equation (2) generates the following competitive equilibrium prices core cost allocation $(x_1, \ldots, x_n)$: $x_i = \phi_i(z_i^*) - (z_i^* - \lambda_i)$, $1 \le i \le n$, where for the open lines

$$x_i = \left(\frac{\alpha_i \bar{z}_i}{\mu_i - \bar{z}_i} + \delta_i\right) - (\bar{z}_i + \delta_i - \lambda_i) = \frac{\alpha_i \bar{z}_i}{\mu_i - \bar{z}_i} - (\bar{z}_i - \lambda_i), \quad 1 \le i \le i^o$$

and for the closed lines

$$x_i = \lambda_i, \quad i^o + 1 \le i \le n.$$

The proof is terminated by substituting $\bar{z}_i, 1 \le i \le i^o$, (see equation (10) into this last expression. □

The competitive equilibrium prices core allocation specified in Theorem 2 demonstrates that lines that are closed at optimality are charged a cost that is proportional to their demand rate, increasing at a rate that is equal to the marginal cost of processing one additional unit in the system. In fact, this core cost allocation is independent of the demand rates and of the congestion cost rates of lines that are closed, which means that in some sense the supervisors of the closed lines are free riders as they are not punished for having inefficient lines. This is even more salient under the case that the outsourcing option is not used, as the demands of the closed lines are processed by the open lines, and the closed lines pay for their total demand, the rate of processing the last infinitesimal unit on the open lines. In addition, we would like to highlight the fact that the structure of the competitive equilibrium prices core allocation for this line balancing game is quite complex, and that it is doubtful if one could guess a core allocation without using Condition 2 in section 2 on market games, especially that this game is not concave. See the next example that rules out the possibility of using Condition 1 in section 2 for proving total balancedness and characterizing the whole core.

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

EXAMPLE 1. Let $N = \{1, 2, 3\}$, with $\mu_1 = \mu_2 = 100$, $\lambda_1 = \lambda_2 = 1$, $\mu_3 = 1$, $\lambda_3 = 0.99$ and $\alpha_1 = \alpha_2 = \alpha_3 = \epsilon$, where ϵ is sufficiently small so that outsourcing is not used by any coalition. Let $S = \{1, 3\}$ and $T = \{2, 3\}$. We have here $\tilde{c}(\{1\}) = \tilde{c}(\{2\}) = 0.01$, $\tilde{c}(\{3\}) = 99$. In coalition S line 1 is open and likewise line 2 is open in coalition T. Thus, $\tilde{c}(S) = \tilde{c}(T) = 0.02$. In coalition $S \cup T$ lines 1 and 2 are open, each getting half of the total traffic. Hence, $\tilde{c}(S \cup T) = 2(\frac{1+0.495}{100-1-0.495})$. It is easy to see $\tilde{c}(S \cap T) = \tilde{c}(\{3\}) = 99$. Hence, $\tilde{c}(S \cup T) + \tilde{c}(S \cap T) > \tilde{c}(S) + \tilde{c}(T)$, proving that $\tilde{c}(\cdot)$ is not a concave set function.

3.2. The Capacity Sharing and Reduction of Capacity Game

In this subsection, similarly to subsection 3.1, we consider a system $N = \{1, \ldots, n\}$ of parallel $M/M/1$ lines, where line $i \in N$ is associated with a Poisson demand rate $\lambda_i > 0$, an exponential service capacity rate μ_i, where $\lambda_i < \mu_i$, and a unit congestion cost rate of $\alpha_i > 0$. As in subsection 3.1, the long run expected cost of the system in case of no cooperation, is the sum of the congestion costs of the individual lines, namely, $\sum_{i \in N} \frac{\alpha_i \lambda_i}{\mu_i - \lambda_i}$, however here, the demand λ_i of line $i \in N$ must be processed by its dedicated line, that is, line i.

We start by considering basic line balancing policies that reassign the whole surplus capacity of $\mu(N) - \lambda(N)$ units to the lines such that each line $i \in N$ gets a positive share of the surplus capacity at top of the necessary minimal level of λ_i. The cost of such a policy is the long run expected congestion cost of the resulting system. However, in general, it might be cheaper for the system to reduce the level of surplus capacity that is used by the lines as a too high surplus capacity might be expensive due to high maintenance costs. In such a case, the capacity that is not used for internal purposes might be left unused in return to some maintenance cost savings or it might be rented out to other firms in return to some income. The cost of the system, in the general case, consists of the congestion cost of the lines in N minus the savings or income due to reducing/renting the extra capacity.

The basic policy that minimizes the total long run congestion cost for the capacity sharing problem (with no option of capacity reduction), where the surplus capacity of $\mu(N) - \lambda(N)$ units is fully allocated to the lines of N, for the special case where the congestion cost rate is the same for all lines, that is, $\alpha_i = \alpha$ for $i \in N$, is derived in Kleinrock (1976, pp. 329–331. We present next a generalization of Kleinrock's solution to line dependent congestion cost parameters. In order to solve this problem, let the function $f_i(s) : \Re_+ \to \Re_+$ denote the long run expected congestion cost of line $i \in N$ if it is allocated $s > 0$ units of surplus capacity:

$$f_i(s) = \frac{\alpha_i \lambda_i}{s}, \quad i \in N. \tag{14}$$

The optimal long run congestion cost is given by

$$c(N) = \min\Big\{ \sum_{i \in N} f_i(s_i) : \sum_{i \in N} s_i = \mu(N) - \lambda(N) \text{ and } s_i > 0 \text{ for } i \in N \Big\}.$$

Basic algebra reveals that the optimal allocation of the surplus capacity to the lines is given by

$$s_i^* = \sqrt{\alpha_i \lambda_i} \cdot \frac{(\mu(N) - \lambda(N))}{\sum_{j \in N} \sqrt{\alpha_j \lambda_j}}, \quad i \in N, \tag{15}$$

and the optimal congestion cost is

$$c(N) = \frac{\left(\sum_{j \in N} \sqrt{\alpha_j \lambda_j}\right)^2}{\mu(N) - \lambda(N)}. \tag{16}$$

Next, we allow for surplus capacity reduction, which comes with unavoidable extra congestion, in return to saving some maintenance costs or, alternatively, earning some rental fees. The savings/income rate per unit of capacity that is not used by the lines is scaled to 1. A capacity reconfiguration of the system allows for a partial reduction of the surplus capacity as well as a reassignment of the remaining surplus capacity among the lines. The total cost consists of the congestion cost minus the savings/income due to capacity's reduction.

Let the function $\phi_i(s)$, defined below, denote the optimal cost of line $i \in N$ whose current surplus capacity is s in view of the option to cut some of its capacity. Recall that the function f is defined in equation (14).

$$\phi_i(s) = \min\{f_i(w) - (s - w) |\ 0 < w \le s\} \quad i \in N. \tag{17}$$

Let $\bar{s}_i$ be the surplus capacity for which the derivative $\frac{\partial f_i(s)}{\partial s} = -\frac{\alpha_i \lambda_i}{s^2}$ equals -1, implying that

$$\bar{s}_i = \sqrt{\alpha_i \lambda_i}, \quad i \in N. \tag{18}$$

Thus, $\bar{s}_i$ is the maximum value of surplus capacity that line i utilizes, as otherwise reducing the capacity is more profitable. Therefore,

$$\phi_i(s) = \begin{cases} \frac{\alpha_i \lambda_i}{s} & \text{if } 0 < s \le \bar{s}_i, \quad i \in N \\ \frac{\alpha_i \lambda_i}{\bar{s}_i} - (s - \bar{s}_i) & \text{otherwise.} \end{cases} \tag{19}$$

We now proceed to the definition of the respective game $G = (N, \tilde{c})$, where $N = \{1, \ldots, n\}$ is the set of $M/M/1$ lines as in subsection 3.1, and the characteristic function $\tilde{c}$ assigns to each coalition $\varnothing \subseteq S \subseteq N$ the

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

optimal congestion cost of its lines minus the savings obtained by reducing its capacity over all feasible policies that assign at most $\mu(S) - \lambda(S)$ units of the surplus capacity among the lines of S, and the rest is reduced. By using the ϕ_i, $i \in N$, functions defined in equation (19), the characteristic function for any coalition $\varnothing \subseteq S \subseteq N$, is expressed by

$$\tilde{c}(S) = \min\Big\{ \sum_{i\in S} \phi_i(s_i) : \sum_{i\in S} s_i = \mu(S) - \lambda(S) \text{ and } s_i > 0 \text{ for } i \in S\Big\}. \quad (20)$$

The cost of the grand coalition, $\tilde{c}(N)$, is obtained by substituting S by N in equation (20).

THEOREM 3. *The capacity sharing with capacity reduction game, $G = (N, \tilde{c})$, where the characteristic function $\tilde{c}$ is defined in equation* (20)*, is a market game.*

PROOF. The proof follows along the same lines as the proof of Theorem 1, using the convexity of $\phi_i(s)$ that follows from the convexity of $f_i(s), i \in N$, see equations (17), and (14). □

In view of Theorem 3 and Condition 2, the capacity sharing with capacity reduction game, $G = (N, \tilde{c})$, is totally balanced and the cost allocation based on competitive equilibrium prices is in its core.

Let Θ be the Lagrange multiplier of the equality constraint in equation (20) for $S = N$. In this problem it holds that $\Theta < 0$ as increasing the surplus capacity of the system reduces the total cost. If the option of capacity reduction is not used, then $\Theta < -1$, and otherwise $\Theta = -1$. More specifically,

$$\theta = \max\left\{-1,\ -\left(\frac{\sum_{i\in N}\sqrt{\alpha_i\lambda_i}}{\mu(N)-\lambda(N)}\right)^2\right\}, \quad (21)$$

where the expression $-\left(\frac{\sum_{i\in N}\sqrt{\alpha_i\lambda_i}}{\mu(N)-\lambda(N)}\right)^2$ is the Lagrange multiplier of the capacity sharing version of the problem where no option of capacity reduction exists. The next Lemma elaborates on the structure of the solution to equation (20) for $S = N$ as follows from the fact that at optimality, the derivatives of $\phi_i(\cdot)$, see equation (19), for $i \in N$, are all identical. The proof is similar to the proof of Lemma 1 so we skip it.

LEMMA 2. *An optimal solution $(s_1^*, \ldots, s_n^*)$ for problem* (20) *for $S = N$, satisfies one of the following two cases:*

- *for all $i \in N$, $s_i^* < \bar{s}_i$ where s_i^* is defined in equation* (15)*, or*
- *for all $i \in N$, $s_i^* \geq \bar{s}_i$.*

Moreover, there exists a unique optimal surplus capacity assignment to lines, where line $i \in N$ is assigned a surplus capacity of $\min\{s_i^*, \bar{s}_i\}$, and $\mu(N) - \sum_{i\in N} \min\{s_i^*, \bar{s}_i\}$ units of surplus capacity are reduced.

The next theorem specifies explicitly the core cost allocation $(x_i)_{i=1}^n$ based on competitive equilibrium prices, see equation (2).

THEOREM 4. *The optimal cost $\tilde{c}(N)$ of the grand coalition and the competitive equilibrium prices cost allocation $(x_i)_{i=1}^n$ of the capacity sharing with surplus capacity reduction game, $(N, \tilde{c})$, are given by:*

$$\tilde{c}(N) = \frac{\left(\sum_{i\in N}\sqrt{\alpha_i\lambda_i}\right)^2}{\mu(N)-\lambda(N)} \quad \text{and} \quad x_i = 2\sqrt{-\Theta}\sqrt{\alpha_i\lambda_i} + \Theta(\mu_i - \lambda_i)\ , \quad i \in N. \quad (22)$$

where Θ is the Lagrange multiplier of the equality constraint in the optimization problem of the grand coalition, see equation (21).

PROOF. Recall the form of the competitive equilibrium prices cost allocation given in equation (2), that is, $x_i = \phi_i(s_i^*) - \Theta(s_i^* - (\mu_i - \lambda_i)), i \in N$, where Θ is the Lagrange multiplier of the equality constraint of the optimization problem of the grand coalition. In the proof, we distinguish between two cases and show that the core cost allocation of both cases boil down to a single form that is based on Θ.

- If $\Theta < -1$, or if $\Theta = -1$ and $\mu(N) - \lambda(N) = \Sigma_{i\in N}\bar{s}_i$, then no reduction of capacity takes place and the surplus capacity $\mu(N) - \lambda(N)$ is distributed among the lines, so that line $i \in N$ is allocated a surplus capacity s_i^* given in equation (15). By equation (21), $\Theta = -\left(\frac{\sum_{i\in N}\sqrt{\alpha_i\lambda_i}}{\mu(N)-\lambda(N)}\right)^2$ and the long run expected congestion cost $\tilde{c}(N)$ is given in equation (16). Substituting these values into the competitive equilibrium prices formula gives $x_i = \frac{\alpha_i\lambda_i}{s_i^*} - \Theta(s_i^*, -(\mu_i - \lambda_i))$, $i \in N$, resulting in the core cost allocation given in equation (22).
- If $\Theta = -1$, and $\mu(N) - \lambda(N) > \Sigma_{i\in N}\bar{s}_i$, then surplus capacity reduction takes place, and line $i \in N$ is assigned a surplus capacity of $\bar{s}_i = \sqrt{\alpha_i\lambda_i}$, see equation (18). Thus, any vector $(s_1^*, \ldots, s_n^*)$ that satisfies $s_i^* = \bar{s}_i + \delta_i$ where $\delta_i > 0$ and $\sum_{i=1}^n s_i^* = \mu(N) - \lambda(N)$, is optimal. However, the surplus capacity assignment $(\bar{s}_i)$, $i \in N$, is unique. The reduction level in capacity is then $\sum_{i\in N}\delta_i = \mu(N) - \lambda(N) - \sum_{i\in N}\bar{s}_i = \mu(N) - \lambda(N) - \sum_{i\in N}\sqrt{\alpha_i\lambda_i} > 0$. The competitive equilibrium prices cost allocation vector is obtained by $x_i = \phi_i(\bar{s}_i + \delta_i) - (-1)(\bar{s}_i + \delta_i - (\mu_i - \lambda_i)) = \phi_i(\bar{s}_i) - \delta_i + (\bar{s}_i +$

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

$\delta_i - (\mu_i - \lambda_i)) = 2\sqrt{\alpha_i\lambda_i} - (\mu_i - \lambda_i), i \in N$, as claimed in equation (22). The optimal cost $\tilde{c}(N)$ is equal to $\sum_{i\in N} x_i$, as the competitive equilibrium prices cost allocation vector $(x_1, \ldots, x_n)$ is in the core, and thus it satisfies the efficiency property. □

In the cost allocation specified in Theorem 4, there are no free riders. The cost allocated to any line $i \in N$ is increasing in its congestion cost rate α_i and in its demand rate λ_i, but is decreasing in the line's contribution capacity of μ_i.

COMMENT 1. *For the game where $\alpha_i = 1$ for all $i \in N$, and with no option of capacity reduction, the vector equation* (22) *is shown in Timmer and Scheinhardt* (2013)*, by a different way, to be in the core.*

The next example shows that the game $G = (N, \tilde{c})$ is not concave.

EXAMPLE 2. Using the queueing system described in Example 1 with $\alpha_i = M$ sufficiently large for all $i \in N$, leading to never opting to save by reducing the capacity under any coalition, results in a game which is not concave: The value of $\tilde{c}(\{1\})$ is large in comparison with the value of any other coalition so concavity is ruled out. Specifically, $\tilde{c}(S) = \tilde{c}(T) = 0.04M$ where $\tilde{c}(S \cap T) = \tilde{c}(\{3\}) = 99M$ and $\tilde{c}(S \cup T) > 0$. Hence, $\tilde{c}(S \cup T) + \tilde{c}(S \cap T) > \tilde{c}(S) + \tilde{c}(T)$, showing that $\tilde{c}$ is not a concave set function.

4. Line Balancing of Parallel Loss Lines

In section 3, the unlimited buffers' size case, is considered. In practice, however, it is quite common that lines have finite buffers that are likely to cause blocking of demand units that arrive to a line when its buffer is full. Blocked units are assumed to be lost forever. A line with a positive finite buffer size may cause two types of cost: (1) the cost of units that reach the line when its buffer is full and therefore are lost, and (2) the cost of units due to having to wait in the buffer. In section 5 we elaborate on possible future research directions on parallel line systems with general finite buffers. We note that from a mathematical point of view, the long run expected cost of lost units in a line with a finite buffer is a function of the respective *loss probability*, that is, the probability that the buffer is full. In this section we focus on the simplest form of the loss probability that applies to a line with no buffer, that is, a buffer of size zero, as a first step toward a future analysis of general, possibly line dependent, buffer sizes. In the next two subsections we consider parallel $M/M/1/1$ line systems where the lines have no buffers, and therefore the cost associated with each line is the cost of its lost units. Note that systems that consist of parallel lines where each demand unit is directed to a line that is idle, if such one exists, and there is no space for waiting units, are called *loss systems*. Thus, a systems of parallel $M/M/1/1$ lines is also called a system of parallel loss lines. In this section, the line balancing techniques described in section 3 are applied and analyzed on parallel $M/M/1/1$ line systems where the cost of lost units replaces the congestion cost. The corresponding cooperative games are shown to be reducible to market games, proving that they are totally balanced. A competitive equilibrium prices cost allocation is suggested for each of these games. It is interesting to note that in contrary section 3, in parallel loss lines, all lines are open in the optimal solution to the unobservable routing problem, where, some lines might be closed in the optimal solution to the capacity sharing problem.

For simplicity, we use the same notation as in section 3 and follow the same assumptions, except that the assumption $\lambda_i < \mu_i$ is not necessary anymore. As the lines have no buffers, the demand that arrives to line $i \in N$ when it is busy processing another unit, is immediately discarded. Let β_i be the cost of a unit lost by line $i \in N$, implying that its long run average cost of lost units is $\frac{\beta_i\lambda_i^2}{\mu_i+\lambda_i}$. The total cost of the system is assumed to be additive in the cost of the individual lines.

4.1. The Unobservable Routing and Outsourcing Game

The optimal domestic unobservable routing policy in a parallel loss line system minimizes the expected average cost of lost units. Each line $i \in N$ is associated with a cost $\beta_i > 0$ per unit lost by the line, an arrival rate $\lambda_i > 0$ and a capacity $\mu_i > 0$. Thus, $\frac{\lambda_i}{\lambda_i+\mu_i}$ is the loss probability, $\frac{\lambda_i^2}{\lambda_i+\mu_i}$ is the loss rate of line $i \in N$, and the total long run expected cost of lost units of the system is $\sum_{i\in N} \frac{\lambda_i^2}{\lambda_i+\mu_i}$. As we show below, it turns out that for line-dependent lost unit cost parameters, no closed form solution of the unobservable routing problem exists. Thus, we first describe the solution for the case where the lost unit cost parameters are identical for all lines in N, that is, $\beta_i = \beta > 0$ for $i \in N$, and then we propose a solution method for line-dependent costs too.

Under the assumption $\beta_i = \beta > 0$ for all $i \in N$, the cost of lost units by line i that faces a demand rate of z, is $\tau_i(z) = \frac{\beta z^2}{\mu_i+z}$. The unobservable routing problem form for any coalition $\varnothing \subseteq S \subseteq N$ of parallel lines is:

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

$$c(S) = \min\Big\{ \sum_{i\in S} \tau_i(z_i) : \sum_{i\in S} z_i = \lambda(S) \quad \text{and } z_i \geq 0 \text{ for } i \in S\Big\}. \tag{23}$$

The unobservable routing problem of the grand coalition is obtained by substituting S by N in equation (23). All lines are open at optimality as $\lim_{z\to 0} \frac{\partial \tau_i(z)}{\partial z} = 0$ for any $\mu_i > 0$ and $\beta > 0$. We define a game $G = (N, c)$ where the set of players N is the set of lines described above, and the characteristic function value for each coalition $\varnothing \subseteq S \subseteq N$ is defined in equation (23).

Let Ψ be the Lagrange multiplier of the equality constraint in equation (23) for $S = N$. Solving equation (23) for $S = N$, by using the KKT conditions, result in

$$\Psi = \beta\left(1 - \left(\frac{\mu(N)}{\mu(N) + \lambda(N)}\right)^2\right). \tag{24}$$

Clearly, $0 < \Psi < \beta$, as the chance of a unit of demand to be lost is less than 1. In particular, the optimal routing to each line is proportional to its capacity:

$$z_i^* = \lambda(N)\frac{\mu_i}{\mu(N)}, \quad i \in N. \tag{25}$$

In addition, at optimality, all lines share the same fraction of busy time, which is equivalent to same chance of a demand unit to be lost :

$$\frac{\lambda(N)}{\lambda(N) + \mu(N)}.$$

The cost of the grand coalition $c(N)$, defined by equation (23) for $S = N$, is calculated by substituting z_i^* for $i \in N$, given in equation (25), into $\sum_{i\in N} \tau_i(z_i)$. The cost of any coalition $\varnothing \subseteq S \subset N$ is given by

$$c(S) = \frac{\beta\lambda(S)^2}{\lambda(S) + \mu(S)}. \tag{26}$$

Interestingly, $c(S)$ is a function of μ_i, $i \in S$, only through the sum $\mu(S)$, though each line works individually.

Next, we consider the respective cooperative game (N, c), defined by the set of lines N, where the characteristic function value of each coalition of lines $\varnothing \subseteq S \subseteq N$ is the expected long run cost of units lost by the lines of S.

THEOREM 5. *The unobservable routing parallel M/M/1/1 lines game (N, c), with the characteristic function defined in equation* (26)*, is a market game. The competitive equilibrium prices cost allocation for this game is given by*

$$x_i = \frac{\beta\lambda(N)}{(\lambda(N)+\mu(N))^2}[(\lambda(N)+2\mu(N))\lambda_i - \lambda(N)\mu_i], \quad i \in N.$$

PROOF. The game (N, c) satisfies the requirements of a market game due to the form of the characteristic function, see equation (23), and to the fact that functions $\tau_i(\cdot)$ in equation (23) are convex for $i \in N$. Therefore, the game is totally balanced. The competitive equilibrium prices is obtained by substituting z_i^* for $i \in N$, see equation (25), and Ψ, see equations (24), in (2). □

As can be seen from the core cost allocation of the game as stated in Theorem 5, the cost allocated to any line i, $i \in N$, is linearly increasing in its demand rate λ_i and linearly decreasing in its capacity μ_i, where λ_i has a greater weight in the cost allocation than μ_i has. In fact, by rearranging the coefficients of λ_i and μ_i, we can see that the cost is increasing linearly in the demand rate λ_i and is decreasing in the surplus capacity $\mu_i - \lambda_i$.

The same calculations for non-identical costs per lost unit by the lines $i \in N$, β_i, reveal that the optimal routing to line i, $i \in N$, is a function of the respective Lagrange multiplier Ψ, that depend on $(\beta_1, \ldots, \beta_n)$. Thus, let $z_i(\Psi)$ be the optimal routing to line $i \in N$ as a function of the respective Lagrange multiplier Ψ. Some basic algebra results in the following equations that have no closed form solutions:

$$z_i(\Psi) = \mu_i\left(\sqrt{1 + \frac{\Psi}{\beta_i - \Psi}} - 1\right), \quad i \in N. \tag{27}$$

The next proposition provides some insight to the line-dependent cost parameters optimization problem and the respective cooperative game:

PROPOSITION 2. *The unobservable routing optimization problem and the respective cooperative game in parallel line loss systems with line dependent cost per unit lost, satisfies the following properties:*

1. *There exists a unique optimal routing of demands to the lines.*
2. *All lines of N are open at optimality.*
3. *No closed form expression for the optimal routing of demands to line exists.*
4. *The game is a market game and therefore it is totally balanced.*

PROOF. Any optimal solution of the unobservable routing in parallel line loss systems with line dependent cost per unit lost, satisfies equation (27). Let $\beta_{min} = \min\{\beta_i : i \in N\}$.

1. The uniqueness of the solution follows from the strict convexity of $\tau_i(z), i \in N$, that implies the uniqueness of the Lagrange multiplier Ψ.

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

2. We first show that $\Psi < \beta_{min}$. For that sake note that the cost incurred by any unit of demand that can be directed to any line, is strictly less than β_{min}, as if this demand unit is routed to the line that is associated with β_{min}, there is a strictly positive chance that it will not get lost, even if all the demand of rate $\lambda(N)$ is routed to that same line. Thus, by equation (27), $z_i(\Psi) > 0$ for $i \in N$.
3. In order to find the optimal routing, one needs to solve for Ψ by solving the constraint $\sum_{i\in N} z_i(\Psi) = \lambda(N)$. This equation has no closed form solution for general $\beta_i, i \in N$.
4. The proof that the game is a market game and is totally balanced follows along the same lines as the proof of Proposition 2. □

As stated in the third item of Proposition 2, in the general case of line dependent lost costs, no closed form solution for Ψ, and therefore neither for $z_i^*(\Psi)$, $i \in N$, exist. However, as $z_i(\Psi)$, $i \in N$, are increasing functions of Ψ, one can search for Ψ by using a technique such as bisection. Once that Ψ, and hence $z_i^*(\Psi)$, $i \in N$, are found, a core cost allocation based on competitive equilibrium prices can be computed.

In the rest of this subsection we limit ourselves to lines that have identical unit lost cost parameters denoted by $\beta > 0$, as done at the beginning of this subsection. In addition, we allow now for outsourcing demand at a cost of 1 per unit. The total cost consists of the cost of lost units plus the outsourcing cost. Assuming that each line is assigned its own dedicated supervisor, the supervisors might be asked to cooperate by redirecting the incoming demand to the lines and possibly to the external service provider in order to minimize the steady state expected cost. Let $G = (N, \tilde{c})$ be the cooperative game that assigns to each coalition of lines $\varnothing \subseteq S \subseteq N$, with given capacities μ_i, $i \in S$, the minimum long run expected cost related to a total demand rate of $\lambda(S)$ units, which are either outsourced or redirected to the lines of S.

If $\beta \leq 1$ then the outsourcing cost of a unit is at least as large as discarding the unit, implying that no unit is outsourced, and the game $G = (N, \tilde{c})$ boils down to the game (N, c) analyzed at the beginning of this subsection, where no option of outsourcing existed. Thus, in the sequel we then consider the case where $\beta > 1$.

The marginal cost due to lost units increases in the demand rate routed to the line. Let $\bar{z}_i$ be the maximum demand rate routed to line i, $i \in N$, before the outsourcing cost is cheaper than the expected loss cost on the line. In order to compute $\bar{z}_i$ we solve for $\frac{\partial \tau_i(z)}{\partial z} = 1$, where $\tau_i(z) = \frac{\beta z^2}{\mu_i + z}$. Therefore, under the case that $\beta > 1$,

$$\bar{z}_i = \left(\sqrt{\frac{\beta}{\beta - 1}} - 1\right) \mu_i \ , \quad i \in N. \tag{28}$$

Clearly, $\bar{z}_i$ increases with μ_i, $i \in N$. Let $g_i(z)$ be the optimal cost of line i for demand rate z:

$$g_i(z) = \begin{cases} \beta z^2/(\mu_i + z) & \text{if } z \leq \bar{z}_i, \quad i \in N \\ \beta \bar{z}_i^2/(\mu_i + \bar{z}_i) + (z - \bar{z}_i) & \text{otherwise.} \end{cases}$$

Thus, the long run expected cost of lost units and outsourcing of any coalition $\varnothing \subseteq S \subseteq N$ in the game $G = (N, \tilde{c})$ is given by

$$\begin{aligned} \tilde{c}(S) = \min\{&\sum_{i\in S} g_i(z_i) : \sum_{i\in S} z_i = \lambda(S) \\ &\text{and } z_i \geq 0 \text{ for } i \in S\}. \end{aligned} \tag{29}$$

The cost of the grand coalition $\tilde{c}(N)$, is obtained by substituting S by N in equation (29). Similarly to equation (12), the Lagrange multiplier corresponding to the equality constraint of problem $\tilde{c}(N)$, see equation (29) for $S = N$, is equal to

$$\Theta = \min\{\Psi, \ 1\}, \tag{30}$$

where Ψ is defined in equation (24).

THEOREM 6. *The unobservable routing with outsourcing in parallel M/M/1/1 lines game, $G = (N, \tilde{c})$, is a market game.*

PROOF. The proof is similar to that of Theorems 1 and 3. Based on the form of the characteristic function of the game, see equation (29), and in view of the convexity of the functions g_i, $i \in N$, the game $G = (N, \tilde{c})$ is a market game, see equation (1). □

Based on Theorem 6 and Condition 2 in section 2, the game $G = (N, \tilde{c})$ is totally balanced, and the cost allocation based on competitive equilibrium prices is in its core.

Similarly to Lemmas 1 and 2, the solution $(z_i^*)_{i\in N}$ of $\tilde{c}(N)$, see equation (29), satisfies that either for all $i \in N$, $z_i^* < \bar{z}_i$ or, for all $i \in N$, $z_i^* \geq \bar{z}_i$. Recall that Θ is the Lagrange multiplier of the equality constraint in equation (29) for $S = N$. If $\Theta < 1$ outsourcing is not used by the lines of N and $z_i^* < \bar{z}_i$, for all $i \in N$. Outsourcing is not used also if $z_i^* = \bar{z}_i$, for all $i \in N$. If outsourcing is not used, the routing rate to line i, $i \in N$, is given in equation (25). Outsourcing is used if and only if $\lambda(N) > \sum_{i\in N} \bar{z}_i$. In such a case, $\Theta = 1$, and the optimal routing to line i, $i \in N$, is $\bar{z}_i$ where a rate of $\lambda(N) - \sum_{i\in N} \bar{z}_i$ units is outsourced. In view of the strict convexity of the functions $g_i(z)$ in $z \in (0, \bar{z}_i)$,

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

for $i \in N$, at optimality, the routing rates to the lines of N are unique.

THEOREM 7. *The optimal cost of the grand coalition and the competitive equilibrium prices core allocation* $(x_i)_{i=1}^n$ *of the unobservable routing parallel M/M/1/1 lines with outsourcing game,* $(N, \tilde{c})$, *are:*

- *If* $\Theta < 1$, *or if* $\Theta = 1$ *and* $\lambda(N) = \sum_{i\in N} \bar{z}_i$, *then no outsourcing takes place by the grand coalition* N, $\tilde{c}(N)$ *is given in equation* (26) *for* $S = N$, *and the cost core allocation for* $i \in N$ *is given in Theorem* 5.
- *Otherwise,* $\Theta = 1$ *and* $\lambda(N) - \sum_{i\in N} \bar{z}_i > 0$, *where* $\bar{z}_i, i \in N$, *are defined in equation* (28). *A rate of* $\lambda(N) - \sum_{i\in N} \bar{z}_i$ *units is outsourced by* N, $\tilde{c}(N) = \lambda(N) - \mu(N)(\sqrt{\beta} - \sqrt{\beta - 1})^2$, *and*

$$x_i = \lambda_i - \mu_i\left(\sqrt{\beta} - \sqrt{\beta - 1}\right)^2, \quad i \in N. \tag{31}$$

PROOF. According to Theorem 6, the cost allocation based on competitive equilibrium prices, whose general form is, $x_i = g_i(z_i^*) - \Theta(z_i^* - \lambda_i)$ *for* $i \in N$, is in the core of the game $(N, \tilde{c})$, where $(z_i^*)_{i=1}^n$ is the optimal solution to equation (29) for $S = N$ and Θ is the respective Lagrange multiplier of its equality constraint, see equation (30).

- If $\Theta < 1$, or if $\Theta = 1$ and $\lambda(N) = \sum_{i\in N} \bar{z}_i$, then

$$\lambda(N) \le \sum_{i\in N} \bar{z}_i = \left(\sqrt{\frac{\beta}{\beta - 1}} - 1\right) \mu(N),$$

and outsourcing is not used by N thus the results obtained by analyzing the unobservable routing game (N, c) at the beginning of this subsection, as stated in the Theorem, apply.

- Otherwise, outsourcing is used, the arrival rate to line i is $\bar{z}_i$, $i \in N$, and a rate of $\lambda(N) - \sum_{i\in N} \bar{z}_i > 0$ demand units is outsourced. Accordingly, $\tilde{c}(N)$ is calculated. The cost allocation in equation (31) is based on competitive equilibrium prices and as the game is a market game, see Theorem 6, it is in its core. □

As all lines are open in the optimal solution, and the cost allocation has the same form for all lines, no free riding issue might occur here.

Note that if outsourcing is not used by N, the form of $\tilde{c}(N)$ is the same as the cost of a single line with capacity $\mu(N)$ and demand of $\lambda(N)$, and moreover it is not additive is λ_i or μ_i, $i \in N$. Thus, guessing a cost allocation that satisfies the efficiency constraint, that is, $\sum_{i=1}^n x_i = \tilde{c}(N)$, is not trivial, let alone guessing a cost allocation that satisfies also the $2^n - 2$ stand-alone constraints, one for each coalition. The identification of the game as a market game allows us to easily get a core cost allocation.

We conclude this subsection with an example that shows that the unobservable routing with outsourcing game in loss systems is not concave:

EXAMPLE 3. Consider an instance where $N = \{1, 2, 3\}$ and $\beta_i = 1$ for $i \in N$. Let, $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$, $\mu_1 = 5$, $\mu_2 = 3$ and $\mu_3 = 4$. It is easy to check that outsourcing is not used by N or by any coalition of lines in N. Take $S = \{1, 2\}$ and $T = \{1, 3\}$. Thus, $S \cap T = \{1\}$ and $S \cup T = N$. $\tilde{c}(S \cup T) = \frac{6^2}{18} = 2$, $\tilde{c}(S \cap T) = \frac{1^2}{6} = 0.1667$, $\tilde{c}(S) = \frac{3^2}{11} = 0.818182$, and $\tilde{c}(T) = \frac{4^2}{13} = 1.2307$, thus $\tilde{c}(S \cup T) + \tilde{c}(S \cap T) > \tilde{c}(S) + \tilde{c}(T)$, proving that the game is not concave, see Condition 1 *in section* 2.

4.2. The Capacity Sharing and Reduction of Capacity Game

The last line balancing game that we introduce is the capacity sharing with possible capacity reduction in a system that consists of parallel $M/M/1/1$ lines. The total cost of a coalition in this game is composed of the cost of loss units plus the possible savings due to capacity reduction. As discussed in section 3.2, the reduced capacity saves the associated maintenance expenses and possibly is rented to other firms for some profit. We use the same notation as in section 4.1, but here, we solve the more general case, that of line dependent cost per unit lost. First, we derive the optimal domestic processing policy that minimizes the long run expected average cost of lost units. The cost due to lost units for line $i \in N$ is given by $f_i(y) = \frac{\beta_i \lambda_i^2}{y + \lambda_i}$, where the variable $y \ge 0$ denotes the capacity assigned to line $i \in N$ and β_i is the cost per unit lost at line $i \in N$. The optimal total cost of lost units of any coalition $S\varnothing \subseteq \subseteq N$ over all policies that reassign the capacity $\mu(S)$ to the lines of S where each line $i \in S$ faces a demand rate of λ_i is given by

$$c(S) = \min\{\sum_{i\in S} f_i(y_i) : \sum_{i\in S} y_i = \mu(S) \text{ and } y_i \ge 0 \text{ for } i \in S\}. \tag{32}$$

The cost of the grand coalition $c(N)$ is computed by substituting S by N in equation (32). Let y_i^* be the optimal capacity of line $i \in N$. Under the optimal policy, lines with a low cost per unit lost are not necessarily open as their capacity can be used by other lines whose cost of lost units is more expensive. Without loss of generality, the lines are assumed to be indexed in a non-increasing order of β_i, that is, $\beta_1 \ge \beta_2 \ldots \ge \beta_n$. We prove that the long run average cost of lost units is minimized by

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

opening lines $\{1, \ldots, i^*\}$, where i^* is of the following form:

$$i^* = \min\left\{ i \in N : \ \beta_{i+1} \le \frac{\left(\sum_{j=1}^{i} \lambda_i \sqrt{\beta_i}\right)^2}{\left(\mu(N) + \sum_{j=1}^{i} \lambda_i\right)^2} \right\}. \quad (33)$$

THEOREM 8. *Consider a capacity sharing parallel M/M/1/1 lines system where the objective is to minimize the long run average cost of lost units by allocating the capacity $\mu(N)$ among the lines of N. The optimal capacity allocation is given by $y_i^* = 0$ for $i > i^*$ (where i^* is given in equation* (33)*), and:*

$$y_i^* = \left(\mu(N) + \sum_{k=1}^{i^*} \lambda_k\right) \frac{\lambda_i \sqrt{\beta_i}}{\sum_{k=1}^{i^*} \lambda_k \sqrt{\beta_k}} - \lambda_i \ \text{for } i \le i^* . \quad (34)$$

In addition, the cost of the grand coalition is given by

$$c(N) = \frac{\left(\sum_{k=1}^{i^*} \lambda_k \sqrt{\beta_k}\right)^2}{\mu(N) + \sum_{k=1}^{i^*} \lambda_k} + \sum_{k=i^*+1}^{n} \beta_k \lambda_k \ , \quad (35)$$

and the Lagrange multiplier of the equality constraint of equation (32) for S = N is given by

$$\Psi = -\frac{\left(\sum_{k=1}^{i^*} \lambda_k \sqrt{\beta_k}\right)^2}{\left(\mu(N) + \sum_{k=1}^{i^*} \lambda_k\right)^2}. \quad (36)$$

PROOF. The proof follows by solving the problem $\min \sum_{i \in N} f_i(y_i)$ under the constraint $\sum_{i \in N} y_i = \mu(N)$ and $y_i \ge 0$ for $i \in N$. Suppose that initially all lines of N have zero capacity, and gradually we allocate them capacities. If line i is closed then it costs $\beta_i \lambda_i$ for losing all its demand. If line $i \in N$ gets an infinitesimal capacity, the marginal cost is $\frac{df_i(y_i)}{dy_i}\big|_{y_i=0} = -\frac{\beta_i \lambda_i^2}{(y_i+\lambda_i)^2}\big|_{y_i=0} = -\beta_i < 0$. Clearly, allocating additional capacity to line $i \in N$ is profitable as fewer demand units of the line are lost. The more capacity is assigned to line $i \in N$, the lower is its marginal profit from the loss of less units. □

Starting with lines with no capacities, line 1 is the first to be allocated capacity until either, the total capacity $\mu(N)$ runs out or its derivative reaches $-\beta_2$. If the first case occurs, only line 1 is open at optimality and the cost is $f_1(\mu(N)) + \sum_{k=2}^{n} \beta_k \lambda_k$. If the second case occurs, continue to allocate capacities to lines 1 and 2 simultaneously and gradually while keeping $\frac{df_1(y_1)}{dy_1}$ equal to $\frac{df_2(y_2)}{dy_2}$ until either, the total capacity $\mu(N)$ runs out, or the derivatives reach the value $-\beta_3$. We continue this process of allocating capacity until all the capacity $\mu(N)$ is allocated. The explicit solution is obtained by solving the KKT conditions of problem (32) for $S = N$ and identifying the Lagrange multiplier Ψ of the equality constraint $\sum_{i \in N} y_i = \mu(N)$. In fact, Ψ is equal to the final value of the derivative of all open lines, that is, the optimal capacity assignment should be such that for each open line i, $\frac{d}{dy_i}\left(\frac{\beta_i \lambda_i^2}{y_i+\lambda_i}\right)\big|_{y_i=y_i^*} = \Psi$ and for each closed line i, the derivative at $y_i^* = 0$ is $-\beta_i \ge \Psi$. By solving the conditions, i^* defined in equation (33), returns the last open line, and for each open line i, $y_i^* = \frac{\sqrt{\beta_i}\lambda_i}{\sqrt{-\Psi}} - \lambda_i$, where Ψ is given in equation (36). As for each open line i, $-\beta_i < \Psi$, the capacity allocated to line i, namely y_i^*, is positive. By substituting the value of Ψ, we get y_i^*, see equation (34), and by substituting y_i^* for $i \in N$, into the cost function $\sum_{i \in N} f_i(y_i^*)$, we get $c(N)$, see equation (35).

If $\beta_i = \beta$ for all $i \in N$, then all lines are open, and each line is allocated a capacity that is proportional to its demand rate, that is, $y_i^* = \mu(N)\frac{\lambda_i}{\lambda(N)}$. In this case, the proportion of time that each line is busy, namely, $\frac{\lambda(N)}{\lambda(N)+\mu(N)}$, and the optimal cost rate, $\beta\frac{\lambda^2(N)}{\lambda(N)+\mu(N)}$, coincide with the corresponding terms in the solution of the unobservable routing in parallel $M/M/1/1/$ lines, analyzed in subsection 4.1.

Next we consider a system of parallel $M/M/1/1$ lines with the option of capacity reduction in return for savings of 1 per unit of capacity reduced. Let $\bar{y}_i$ be the maximum capacity allocated to line $i \in N$ before the capacity reduction option becomes more profitable. A line $i \in N$ is associated with a positive $\bar{y}_i$ if and only if $\frac{df_i(y_i)}{dy_i}\big|_{y_i=0} = -\beta_i < -1$. Some basic algebra gives

$$\bar{y}_i = \max\{0, \lambda_i(\sqrt{\beta_i} - 1)\}.$$

The cost of line $i \in N$ as a function of its capacity $y \ge 0$ is given by

$$\phi_i(y) = \begin{cases} \beta_i \lambda_i^2/(y + \lambda_i) & \text{if } \ y \le \bar{y}_i \\ \beta_i \lambda_i^2/(\bar{y}_i + \lambda_i) - (y - \bar{y}_i) & \text{otherwise} \end{cases} \quad (37)$$

Let $G = (N, \tilde{c})$ be the respective cooperative game on a parallel $M/M/1/1$ lines system, where line $i \in N$ is associated with a demand rate λ_i, and a total capacity of $\mu(N)$ is allocated to the lines of N, with possible capacity reduction. The respective characteristic function value of the grand coalition given in equation (38) returns the minimum long run expected cost of lost units minus savings due to capacity reduction over all feasible policies:

$$\tilde{c}(S) = \min\left\{ \sum_{i \in S} \phi_i(y_i) : \sum_{i \in S} y_i = \mu(S) \text{ and } y_i \ge 0 \text{ for } i \in S \right\}. \quad (38)$$

The cost of the grand coalition $\tilde{c}(N)$ is found by substituting S by N in equation (38).

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

THEOREM 9. *The capacity sharing with capacity reduction in parallel M/M/1/1 lines game $G = (N, \tilde{c})$, where the characteristic function $\tilde{c}$ is defined equation* (38)*, is a market game.*

PROOF. The Theorem follows directly from the characteristic function, and the convexity of the functions $\phi_i(y)$, $i \in N$, see equation (37). □

In view of Theorem 9 the game is totally balanced, and the cost allocation based on competitive equilibrium prices is in its core. Let Θ be the Lagrange multiplier of the equality constraint in equation (38) for $S = N$. Clearly, $\Theta < 0$ as increasing the capacity of the system may only reduce the total cost. If the option of capacity reduction is not used, then $\Theta = \Psi < -1$, and otherwise $\Theta = -1$. More specifically,

$$\Theta = \min\{-1, \Psi\}, \tag{39}$$

where Ψ is the Lagrange multiplier of the problem with no option of capacity reduction, see equation (36). Clearly, the optimal solution to the problem with the additional option of capacity reduction is such that all lines $i \in N$ with $\beta_i < 1$, are shut as the marginal savings of opening such a line is β_i, which is lower than the marginal revenue of reducing the capacity. In particular, if $\beta_1 \leq 1$ then all lines are shut, all demand is lost, all capacity is reduced, and the long run expected cost is $\sum_{i\in N} \beta_i\lambda_i - \mu(N)$. Otherwise, let $i' = \min\{i \in N : \beta_i \leq 1\} - 1$. Thus, line $i \in N$ is closed if and only if $i > \min\{i^*, i'\} \stackrel{\text{def}}{=} i^0$, where i^* is defined in equation (33). Similarly to the other games described in this study, at optimality, either for all $i \leq i^0, y_i^* < \bar{y}_i$ or, for all $i \leq i^0, y_i^* \geq \bar{y}_i$.

As is demonstrated in Theorem 10, surprisingly, this game has a unique form of the equilibrium competitive prices cost allocation that is independent of whether or not reduction of capacity takes place.

THEOREM 10. *The cost of the grand coalition and the competitive equilibrium prices core allocation $(x_i)_{i=1}^n$ of the capacity sharing parallel M/M/1/1 lines with capacity reduction game $(N, \tilde{c})$, are given by:*

$$\tilde{c}(N) = 2\sqrt{-\Theta}\sum_{i=1}^{i^o}\lambda_i\sqrt{\beta_i} + \Theta\left(\sum_{i=1}^{i^o}\lambda_i + \mu(N)\right) + \sum_{i^o+1}^{n}\beta_i\lambda_i,$$

where $\Theta = \min\{-1, \Psi\}$ with Ψ defined in equation (36).

$$x_i = 2\lambda_i\sqrt{-\Theta\beta_i} + (\lambda_i + \mu_i)\Theta \quad \text{for } i \leq i^o,$$

$$x_i = \lambda_i\beta_i + \Theta\mu_i \quad \text{for } i^o < i \leq n.$$

PROOF. The proof follows directly from the above analysis, the fact that the game is a market game, see Theorem 9, and the form of an competitive equilibrium prices cost allocation, see equation (2). □

We note that in this problem, at optimality, some lines might be open and the others might be closed, as in subsection 3.1. However, here, as we explain below, the competitive equilibrium prices core allocation of line $i \in N$, given in Theorem 10, which is a linear function of both the demand rate λ_i and the service rate μ_i, of the line, does not seem to cause adverse feelings of free riding as we encountered in subsection 3.1. Each line is compensated for all its capacity at the rate of $|\Theta|$, see equation (39). A close line $i \in N$, pays for the loss of all of its demand a rate of β_i per unit, that is, it pays a rate of $\beta_i\lambda_i$. An open line $j \in N$, pays for the loss of demand $\lambda_j(2\sqrt{-\Theta\beta_j} + \Theta) < \beta_j\lambda_j$, meaning that it pays the loss fee for just a fraction of its demand.

As we have shown for the other games, this last game is also not concave. In fact, Example 3 shows that the capacity sharing loss system with capacity reduction game is not concave for the case that $\beta_i = \beta$ for $i = 1, 2, 3$.

5. Conclusion

The area of line balancing is fundamental in operations and service management. It allows a firm to increase its profit and improve its efficiency. It is well known that in practice, various causes may stand as obstacles in achieving a stable full cooperation among various units in a firm. One way that the management may mitigate these obstacles and encourage full cooperation is by displaying a scheme that allocates the total cost or rewards among the cooperating units so that both the strengthes and deficiencies of each unit, are reflected by the scheme. This can be done by using the theory of cooperative games and the various cost allocation concepts that have been proposed in the literature. In this study, we focus on cost allocation schemes that guarantee full cooperation and the stability of the grand coalition, namely, no unit or coalition of units has an incentive to abandon the grand coalition. We consider here four line balancing games that are reducible to market games, and as such we could point out for each game a specific core cost allocation based on competitive equilibrium prices. The competitive equilibrium prices cost allocation

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708

assigns to each player $i \in N$ the cost that player i faces after applying an optimal assignment of the players' initial resources among themselves, minus the economic value of the additional resources that are assigned to player i, where each resource type is evaluated by its corresponding Lagrange multiplier. We note that, in general, market games deal with any number of resources. In this article, all the line balancing games that we consider, have a single resource that is reallocated among the players. In the general case, however, player i may get more units of some resources, where simultaneously, the player may be asked to transfer some (or all of the) units of other resources that she owns to the other players. By definition, this cost allocation sounds fair, and indeed it is proved to be in the core of the game, that is, no coalition $\varnothing \subseteq S \subseteq N$ can claim that it can pay less by abandoning the grand coalition. Still, it does not mean that all players will be satisfied by being charged according to their competitive equilibrium prices. See our discussion in subsection 3.1 about possible adverse feelings on free riding of supervisors of lines that are closed at optimality.

As discussed in section 1, in practice, lines usually have buffers of finite size. In production processes, the buffers are usually finite due to space limitations that result from the cost of the space, the cost of holding units in a buffer, or the opportunity cost of financing WIP within a buffer, see [7]. Similarly, in service systems, the buffers' size are determined by capacity limitations that are the result of either limited space, or of a constrained technology, e.g., the number of IP IVR ports in a call center that are used for queueing purposes. A general parallel line system may have line dependent buffer sizes. The cost of a line with a positive finite-size buffer consists of both the congestion cost and the cost of demand lost. In a general system each line $i \in N$ may be associated with three types of resources (1) its buffer size b_i, where $0 \leq b_i \leq \infty$, (2) its demand rate $\lambda_i \geq 0$ and (3) its service capacity rate $\mu_i \geq 0$. In addition, each such line is associated with two cost parameters (i) a cost $\beta_i \geq 0$ per unit lost, which is applicable if $b_i < \infty$, and (ii) a congestion cost rate $\alpha_i \geq 0$, which is applicable if $b_i > 0$. By considering the redistribution of the at most three resources listed above, there exist seven variants of line balancing games on parallel line systems. The study of such systems can help managers to improve their systems by applying line balancing methods while being sensitive to designing fair cost/bonus allocation schemes that retain the stability of the whole system and the continuing cooperation among the heads of the different units in the firm.

Acknowledgments

The research of the first author has been supported by the Israel Science Foundation, grant no. 109/12 and 338/15 of the author, and also, it has been partially funded by the Israeli Institute for Business Research. The research of the second author was partially funded by the Israel Science Foundation, grant no. 511/15.

References

Altman, E, U. Ayesta, B. Prabha. 2011. Line balancing in processor sharing systems. *Telecommun. Syst.* **47**(1–2): 35–48.

Anily, S. 2017. The characterization of the nonnegative core of a class of centralized aggregated cooperative games. Working paper, IIBR Coller School of Management, Tel Aviv University, Israel.

Anily, S., M. Haviv. 2010. Cooperation in service systems. *Oper. Res.* **58**(3): 660–673.

Anily, S., M. Haviv. 2014. Subadditive homogenous of degree 1 games are totally balanced. *Oper. Res.* **62**(4): 788–793.

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

Chakravarthy, S. R. 2016. Queueing models with optional cooperative services. *Eur. J. Oper. Res.* **248**(3): 997–1008.

Hassin, R., M. Haviv. 2003. *To Queue or Not to Queue: Equilibrium Behavior in Queues*. Kluwer Academic Publishers, Norwell, MA, 2061, USA.

Karsten, F. 2013. Resource Pooling Games. Unpublished doctoral dissertation, Eindhoven University of Technology, the Netherlands.

Karsten, F., M. Slikker, G.-J. van Houtum. 2009. Spare parts inventory pooling games. Beta Working Paper series 300, Eindhoven University of Technology, the Netherlands.

Karsten, F., M. Slikker, G.-J. van Houtum. 2011. Analysis of resource pooling games via a new extenstion of the Erlang loss function. Beta Working Paper series 344, Eindhoven University of Technology, the Netherlands.

Kleinrock, L. 1976. *Queueing Systems, Volume 2: Computer Applications*, John Wiley and Sons Inc, New York.

Osborne, M. J., A. Rubinstein. 1994. *A Course in Game Theory*, The MIT Press, Bighamton, New York.

Peleg, B., P. Sudholter. 2007. *Introduction to the Theory of Cooperative Games*, 2nd edn. Kluwer, Berlin.

Shapley, L. S. 1971. Cores of concave games. *Int. J. Game Theory* **1**: 11–26.

Shapley, L. S., M. Shubik. 1969. On market games. *J. Econ. Theory* **1**: 9–25.

Timmer, J., W. Scheinhardt. 2013. Cost sharing of cooperating queues in a Jackson network. *Queueing Syst. Theory Appl.* **75**(1): 1–17.

Yu, Y., S. Benjaafar, Y. Gerchak. 2015. Capacity sharing and cost allocation among independent firms with congestion. *Prod. Oper. Manag.* **24**(8): 1285–1310.

Please Cite this article in press as: Anily, S., M. Haviv. Line Balancing in Parallel $M/M/1$ Lines and Loss Systems as Cooperative Games. *Production and Operations Management* (2017), doi 10.1111/poms.12708