European Journal of Operational Research 214 (2011) 674–682

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

The impact of transportation delays on repairshop capacity pooling and spare part inventories

Pedram Sahba, Barış Balcıoğlu *

University of Toronto, Department of Mechanical and Industrial Engineering, 5 King's College Rd., Toronto, ON, Canada M5S 3G8

ARTICLE INFO

Article history:
Received 20 July 2010
Accepted 17 May 2011
Available online 23 May 2011

Keywords:
Spare parts
Multiple finite-population queueing systems
Transportation delays
Mean-Value Analysis

ABSTRACT

In this paper, to serve different fleets of machines at different locations, we study whether repair shop pooling is more cost effective than having dedicated on-site repair shops for each fleet. When modeling the former alternative, we take transportation delays and related costs into account and represent it as a closed queueing network. This allows us to include on-site spare-part inventories that operate according to a continuous-review base-stock policy. We obtain the steady-state distribution of components at each location and the cost of the system with a pooled repair shop by applying the Mean-Value Analysis technique. Our numerical findings indicate that when transportation costs are reasonable, repair shop pooling is a better alternative.

1. Introduction

Capacity pooling is an important theme in the queueing and operations management literature. To determine whether capacity pooling is beneficial, a queueing system with pooled service capacity is usually contrasted with a system of m resources serving different streams of independent arrivals. Pooling is conceived in two ways. First, in the *pooling of service rates* independent resources are consolidated with a single server providing a faster service rate (Yu et al., Under Review). In our paper, the service rate of the server with pooled capacity is the sum of the service rates of independent resources. Second, in the *consolidation of servers* multiple servers are placed at a single location (Smith and Whitt, 1981; Benjaafar, 1995). Either way, capacity pooling usually decreases the total system costs (for design problems on choosing the number of servers and service rates, see Stidham, 1970). However, when queueing systems are used to analyze supply chains, production capacity pooling cannot be solely modeled by a faster single server or a larger group of servers at one location. In such problems, capacity pooling implies that the products cannot be delivered to different markets instantaneously and that their delivery costs the decision makers. The same issue arises in inventory pooling, too (Eppen, 1979; Gerchak and He, 2003; Benjaafar et al., 2005). Thus, transportation delays and costs have to be incorporated into models of the pooled system, something that has been ignored in earlier research.

In this paper, we explore the effect of transportation delays and costs on the benefits of capacity pooling in a repair/maintenance shop. The decision maker can be an outsourcing company serving a number of clients or the maintenance department of a company that serves various branches of the parent company. Accordingly, each client or branch at a different location is a fleet of identical machines, and each machine is subject to failure due to a single critical component, e.g., engines. When a component fails, it is sent to a repair shop to be fixed. To reduce down time, a stock of critical components reserved for each fleet is kept as spare parts on-site. If there is stock, a spare component is instantaneously installed on the failed machine. Otherwise, the failed machine is down until a repaired component can be dispatched from the repair shop. If all machines are functional, the repaired component is placed in the spare part inventory. In production/inventory systems, production and deliveries are usually done in batches. In contrast, when components are expensive, and failures are rare (equivalently, times to failures are much longer than repair times) it is assumed that the broken (fixed) components to (from) the repair shop are sent (received) one by one (Graves, 1985; Caggiano et al., 2009). In our problem, we consider continuous-review base-stock policies for controlling spare-part inventories.

In this setting, the decision maker has two alternatives. In the first, a separate on-site repair shop can be dedicated to the fleet at that location. The advantage is that fleets do not suffer from transportation delays and the decision maker (alternatively, the system) does not incur transportation costs. In the second, a centralized repair shop with a higher capacity serves all fleets. Thus, some locations experience transportation delays and the system incurs transportation costs. However, a higher capacity drastically

* Corresponding author. Tel.: +1 416 978 57 46.
E-mail addresses: pedram@mie.utoronto.ca (P. Sahba), baris@mie.utoronto.ca (B. Balcıoğlu).

0377-2217/$ - see front matter Crown Copyright © 2011 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2011.05.022

reduces repair times and can prevent lengthy down times. By comparing these two systems, we address the important question of when repair shop pooling is beneficial if transportation times and costs are not negligible.

The results of this research are important to maintenance outsourcing companies and large companies that operate and maintain manufacturing plants at different locations. In both cases, the goal is decreasing maintenance costs while not increasing production stoppages and losses. For a company, the down time cost includes the cost of production losses and the cost of repair. The holding cost includes capital costs of the investment tied up in stock as well as operational costs of warehousing (Silver et al., 1998, p. 45). Both holding and down time costs are regularly expressed per item per unit time (Louit et al., 2011). Transportation costs refer to costs incurred for handling and delivering broken and fixed components between the repair shop and the fleet locations. In this context, whether to pool repair shop capacity is an important consideration. However, maintenance activities are increasingly contracted out to outsourcing companies (Hui and Tsang, 2004; Kumar and Markeset, 2007), and clients expect the outsourcing company to perform repair services and supply spare parts and logistics, etc. From the outsourcing company's point of view, the down time cost of a fleet may be the penalty cost it will pay the client if machines become down. It is reasonable that the outsourcing company is responsible for deliveries without charging transportation costs to the client. It can also be agreed that inventory or repair facility space should be allocated on-site by the client, as the outsourcing company will be tying up its capital in spare parts and will be operating the inventories and the repair shop(s). Given this, cost minimization is an objective for both an outsourcing company and a maintenance department of the parent company. In both cases, the decision maker needs to compare the pros and cons of repair shop pooling at a single location.

In this context, the nature and the formulation of the problem can make analysis quite difficult. A simple system of one repair shop and one spare parts inventory at each location can be analyzed by a birth-and-death process, e.g., Taylor and Jackson (1954). But if failed components from each fleet are treated to form a separate class of customers, and all customer types are served by a centralized resource, (e.g., the centralized repair shop), the problem turns into a multi-class queueing system. If the repair shop is modeled by an infinite server group, and failure rate at each location is considered constant, assuming deterministic transportation times, the approximation due to Graves (1985) can be used to determine base-stock levels at each location. Similarly, assuming constant failure rates from each fleet and considering an infinite centralized inventory (instead of a centralized repair shop) from which new non-repairable service parts are sent to local warehouses when needed, Kutanoglu and Mahajan (2009) include transportation delays in their model. Our problem, on the other hand, is a queueing system with finite calling populations. In our problem, multiple fleets are served by a single repair shop, making it a machine interference problem (MIP) (see Haque and Armstrong, 2007 for a recent literature survey). Observe that we consider state-dependent customer arrival rates and a single server whereas Graves (1985) considers constant customer arrival rates and an infinite server group. The typical solution for the MIP is to model the underlying queueing system with finite calling populations (Haque and Armstrong, 2007) to obtain the steady-state performance measures. However, multi-class systems such as our problem with state-dependent failure rates (failure rates depend on the number of functional machines in the fleet) served by a centralized repair shop with local spare parts inventories at each location is difficult to analyze, even with a first-come-first-served (FCFS) dispatching policy. Incorporating transportation delays in this model is even more challenging. In fact, even in a production/inventory setting where demand rates are assumed to be constant, incorporating transportation delays in the underlying queueing model is difficult. This may explain why the impact of transportation delays and costs has not been addressed in the literature on resource pooling. Simulation is a viable yet costly approach used by Sahba and Balcıoğlu (Under Review) to assess the impact of transportation delays on certain spare part provisioning problems with a centralized repair shop. The main contribution of this paper is the introduction of a queueing model in an unexplored field; more specifically, we consider the impact of transportation delays and costs on an inventory/repair shop system.

To this end, we model this system as a closed queueing network; instead of balance equations, we exploit the Mean-Value Analysis (MVA) developed by Reiser and Lavenberg (1980) to obtain the stationary system size distribution. MVA, like the convolution algorithm (Buzen, 1973), is a numerical algorithm that takes advantage of the product form property of queueing networks with certain conditions (see Gordon and Newell, 1967). Since closed form cost functions are not available, we perform an extensive numerical study, the results of which, we believe, are important. Repair shop pooling is beneficial when transportation delays and costs are negligible (Sahba and Balcıoğlu, Under Review), but it is not always the case in our problem. However, when transportation costs are not unreasonably high, as one would expect in land transportation, even fleets at long distances can be served from a centralized repair shop; the system will incur less cost than would dedicated on-site repair shops. Moreover, repair shop pooling is more attractive if fleet sizes increase or machines become more unreliable. Our paper demonstrates the benefits of capacity pooling in realistic settings. Since, to the best of our knowledge, this is the first paper to include transportation delays and costs in pooling problems with finite repair capacity, it will trigger interest in production/inventory systems for which land and sea transportation are widely used. An additional consideration in repair/spare part inventory and production/inventory systems would be adding a centralized inventory (for benefits of a centralized inventory backing up on-site inventories, see Sahba and Balcıoğlu, Under Review).

The rest of the paper is organized as follows. In Section 2, we present the two alternatives compared in this paper. In Section 3, we define a repair system network for a centralized repair shop. In Section 4, we present the algorithm to obtain the cost of this system. The results of a numerical study comparing two alternatives is presented in Section 5. Finally, in Section 6, we conclude our study and discuss our future research questions.

2. Two alternative repair systems

We consider a system of m fleets at different locations (e.g., manufacturing plants or mines). Each fleet i has N_i machines (interchangeably referred to as type i machine), $i = 1,\ldots,m$, and aims to have all machines functional at all times to continue production at targeted levels. Each machine is subject to failure due to a single repairable component which is not necessarily the same across all locations but allowed to be location-specific. We assume that times to failure for each machine/component follow an independent exponential distribution with possibly different rates, λ_i. When a machine fails, the broken component is repaired at a designated repair shop. In this case, the repair shop is modeled as a single server queueing system with exponential service/repair times that do not depend on the origin'' of the failed component. To reduce the unavailability of a failed machine, S_i units of the critical component are kept in a spare parts inventory at location i (namely, inventory i) for fleet i. Thus, the failed component can be replaced immediately if a spare part is available, thereby