European Journal of Operational Research 196 (2009) 162–170

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Production, Manufacturing and Logistics

Workforce scheduling with multiple objectives

Ignacio Castillo [a], Tarja Joro [b,*], Yong Yue Li [c]

[a] School of Business and Economics, Wilfrid Laurier University, Waterloo, ON, Canada N2L 3C5
[b] University of Alberta School of Business, Department of Finance and Management Science, Edmonton, AB, Canada T6G 2R6
[c] Frontline Systems Inc., P.O. Box 4288, Incline Village, NV 89451, USA

ARTICLE INFO

Article history:
Received 17 December 2005
Accepted 25 February 2008
Available online 18 March 2008

Keywords:
Service operations
Personnel and shift scheduling
Multiple criteria

ABSTRACT

In workforce scheduling, the optimal schedule has traditionally been determined by minimizing the cost of labor subject to an acceptable service level, which is defined as the percentage of customers served within a predetermined time interval. We propose an alternative multidimensional paradigm, where cost minimization and service level maximization are considered simultaneously, together with other, complementary criteria. The ultimate goal of the proposed approach is to open a broader workforce scheduling paradigm that incorporates service quality into the analysis and provides the possibility to study the interaction between cost and service quality. Furthermore, the approach enables us to avoid strong assumptions. An example with real-world, empirical demand data is provided.

1. Introduction

Workforce scheduling aims to find employee shift arrangements to match a time-varying customer demand for service while keeping costs under control and satisfying all applicable regulations (e.g., shift lengths and spacing of breaks). The current paradigm in workforce scheduling is to minimize the cost of labor subject to a target service level, which is defined as the percentage of customers served within a predetermined time interval. Since the seminal papers by Edie (1954) and Dantzig (1954), this paradigm has been applied to workforce scheduling in many organizations – from police departments to laboratories to call centers (see Agnihothri and Taylor, 1991; Brusco et al., 1995; Callahan and Khan, 1993; Gopalakrishnan et al., 1993; Harris et al., 1987; Sze, 1984; and Taylor and Huxley, 1989, among others).

While the use of a target service level can currently be considered as a sector-wide standard, one can argue that a single operational measure is not sufficient to capture the performance of service organizations nor to characterize and quantify service quality. Even though the relationship between customer waiting and service quality has been discussed widely (see Davis, 1991; Davis and Heineke, 1994; Davis and Maggard, 1990; and Taylor, 1994, for example), there is little evidence in the service-sector research literature that suggests what operational criteria are related to service quality.

This observation is not new: indeed, many other researchers have previously acknowledged the ambiguous relationship between a target service level and service quality, and have attempted different approaches to resolve or avoid the ambiguity. One of the approaches is to expand the cost component of the objective function to account for the cost of poor service and the cost of waiting (see Andrews and Parsons, 1993; Easton and Goodale, 2005; Goodale et al., 2003; Grassmann, 1988; Koelling and Bailey, 1984; Mabert, 1979, for example). The challenge in this approach, however, is the estimation of these costs (see Baker, 1976; Taha, 1981, for example). Such estimation is likely to be unique for each type of service organization because it depends on the customers' response to waiting, which is affected by various factors that differ from one organization to another (see Jackson, 2002; Katz et al., 1991, for example). It is clear, then, that this approach has not offered a definitive answer about to how to capture service quality, and that there is interest in the development of alternative approaches.

In addition to the difficulties in modeling the service quality, another concern in traditional workforce scheduling is the use of some rather limiting assumptions. These include the use of exponentially distributed service times resulting from the use of $M/M/s$ queueing models, the steady state assumption arising from the traditional Stationary Independent Period by Period approach, and the sequential nature of the steps taken to determine the server requirements and actual schedule.

The aim of this paper is to open a new workforce-scheduling paradigm that avoids the above-mentioned assumptions and provides a more comprehensive view of service quality. Instead of aiming to find a single optimal solution, the proposed approach generates a large number of plausible schedules to be evaluated with multiple criteria. Among the evaluated schedules, the efficient ones are identified and, among those, the best one is chosen. The novelty of the approach is that schedules are generated directly from the demand

* Corresponding author.
E-mail addresses: icastillo@wlu.ca (I. Castillo), tarja.joro@ualberta.ca (T. Joro), yong@solver.com (Y.Y. Li).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.02.038

profile and are later evaluated using a tool that does not necessitate the limiting assumptions in terms of the actual operational characteristics of the organization. Moreover, the evaluation of several different criteria gives the management a broader understanding of service quality and its relationship to labor expenditures.

Furthermore, the proposed paradigm enables one to characterize and quantify of service quality using multiple, complementary criteria. These criteria can consist of (but are not limited to) average waiting time and queue length, upper tail measures for waiting time and queue length (such as maximum or 95th percentile), abandonment and blocking rates, service levels with various threshold times, and personnel utilization. The list can be adjusted based on the needs of the organization under consideration.

The consideration of these complementary criteria aims at resolving the service-quality ambiguity since neither a single operational measure nor a questionable cost term is used. When identifying the efficient schedules and choosing the best one, all these criteria are considered simultaneously. In addition, the interactive selection procedure allows management to investigate the relationship between cost and various service-quality criteria, rather than simply forcing managers to attempt the execution of the minimum-cost schedule that achieves a target service level.

In Section 2, we discuss the limitations of the traditional approach. We describe the proposed paradigm in detail in Section 3. Section 4 provides an example using real-world, empirical demand data from the call center of a major North American utility company. The conclusion and directions for further research are provided in Section 5.

2. Traditional approach

The traditional approach of workforce scheduling consists of four steps (Thompson, 1995): (1) forecast demand; (2) obtain staff requirements based on the forecasted demand; (3) schedule shifts to meet staff requirements; and (4) real-time control. Since *Steps 1 and 4* fall outside of the actual scheduling process, our emphasis is on *Steps 2 and 3*.

In *Step 2*, an $M/M/s$ queueing model is commonly used to estimate the stationary system performance. The stationary system performance requires the average arrival and service rates to be constant throughout the day. However, in real life, most organizations have time-varying demand. In order to use an $M/M/s$ queueing model under this condition, the service period is partitioned into equal, short planning periods (usually 30 minutes or 1 hour). It is assumed that the average arrival and service rates are constant in each planning period and that the system reaches steady state at the beginning of each planning period. With these assumptions, an $M/M/s$ queueing model can then be used to compute the minimum number of customer representatives needed to meet a certain service level in each planning period. Note that each planning period is treated independently in this approach. This approach of obtaining staffing requirements (the minimum number of representatives needed to meet a target service level in each planning period) is usually called the Stationary Independent Period by Period (SIPP) approach.

One of the limiting assumptions is that the service time follows an exponential distribution: This is rarely the case in the real-world. Furthermore, note that abandonment and blocking behaviors that are well documented in the service-sector research literature are not considered due to the simplified $M/M/s$ modeling assumptions. However, the critical problem of the SIPP approach is that it assumes that each planning period is independent; therefore, and especially in heavy traffic periods, the customers that are waiting in the previous period will not be carried over to the next period, likely providing understaffed results for the latter periods. This limitation was first identified in Baker (1976). As the planning periods get shorter, the stationary state assumption is also questionable in the SIPP approach. The reader is also referred to Thompson (1995), Easton and Rossin (1996), Goodale et al. (2003) and Easton and Goodale (2005).

Step 3 of the traditional approach transforms daily staffing requirements into a schedule. A shift is defined as a set of intervals in a day during which a customer representative works, and a schedule here refers to a set of shifts that provides the total staffing requirement for a day. For example, a feasible shift may require a customer representative to work from 8:30 to 17:30, with a half-hour lunch break at 12:00 and two 15-min coffee breaks, one at 10:00 and the other at 15:30. The scheduling problem is an optimization program that minimizes the cost of labor of a schedule while satisfying the staffing requirements in each planning period. There is a considerable amount of research literature available that attempts to solve this problem through various approaches with various perspectives (see Gans et al., 2003, for a brief summary). The traditional approach is to use an integer linear programming program to choose from all feasible shifts to cover the staffing requirements at the minimum cost. Due to the large number of possible shifts, the optimization program is complicated to formulate and computationally expensive to solve: an optimal solution cannot usually be guaranteed.

More importantly, even if an optimal solution is obtained, the schedule does not necessarily provide the target service level established in the first place due to the intrinsic deficiencies in *Steps 2 and 3*. Recall that in *Step 2*, several limiting assumptions were made to facilitate the use of the SIPP approach. Green et al. (2001) identify the conditions under which the SIPP approach fails to provide the target service level and highlight the underlying reasons.

Another apparent limitation lies in the sequential nature of *Steps 2 and 3*. Several studies indicate that workforce scheduling in sequence from *Step 2* to *Step 3* can give misleading results (for details see Easton and Rossin, 1996 and Thompson, 1999). The main drawback presented in this sequence is that the optimal schedule is generated without considering the employee information; as a result, employees scheduled might not be available or the schedules produced might violate the applicable regulations (e.g., shift lengths and spacing of breaks). Tien and Kamiyama (1982) indicate that further research is needed to simultaneously consider *Steps 2 to 3*. Moreover, recent studies provide evidence that significant cost savings can be obtained if these steps are considered simultaneously (see Ingolfsson et al., 2002; Ingolfsson and Cabral, 2002; Cezik and L'Ecuyer, 2005; Green et al., 2001, for example). Thus, integrating *Steps 2 and 3* is arguably the preferred way to approach the scheduling problem.

In the proposed paradigm, the schedules are generated directly from the demand profile; thus, separate *Steps 2 and 3* are not needed. Overall, the proposed approach opens a broader workforce scheduling paradigm that enables us to consider multiple, complimentary criteria for service quality simultaneously as well as study their interactions with labor cost.

3. Proposed paradigm

We present in this section the proposed paradigm for obtaining the best schedule accounting for multiple, complimentary criteria. The process of the new workforce-scheduling paradigm is:

(1) identify all possible shifts,
(2) create demand profiles,
(3) use demand profiles to generate plausible schedules with various combinations of the possible shifts,
(4) obtain performance criteria for the schedules, and