European Journal of Operational Research 236 (2014) 14–26

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

Dynamic resource allocation: A flexible and tractable modeling framework

Dimitris Bertsimas [a,*], Shubham Gupta [b], Guglielmo Lulli [c]

[a] Operations Research Center, Massachusetts Institute of Technology, E40-147, Cambridge, MA 02139-4307, United States
[b] Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, United States
[c] Dept. of Informatics, Systems and Communication, University of Milano-Bicocca, 20126 Milano, Italy

ARTICLE INFO

Article history:
Received 7 December 2012
Accepted 27 October 2013
Available online 14 November 2013

Keywords:
Applications of integer optimization
Resource allocation
Scheduling
Fairness

ABSTRACT

This paper presents a binary optimization framework for modeling dynamic resource allocation problems. The framework (a) allows modeling flexibility by incorporating different objective functions, alternative sets of resources and fairness controls; (b) is widely applicable in a variety of problems in transportation, services and engineering; and (c) is tractable, i.e., provides near optimal solutions fast for large-scale instances. To justify these assertions, we model and report encouraging computational results on three widely studied problems – the Air Traffic Flow Management, the Aircraft Maintenance Problems and Job Shop Scheduling. Finally, we provide several polyhedral results that offer insights on its effectiveness.

1. Introduction

Allocation of resources over time is a problem of significant importance that many organizations in industry, government and education face. Correspondingly, resource allocation problems have received considerable attention in the Operations Research literature. In real world applications of resource allocation problems, specific issues arise: assignment of requests to resources over time, allowing the flexibility of utilizing alternative resources to complete the requests, fairness issues among different requests, among others. There has been extensive work on specific examples of resource allocation problems (for example, the extensive literature on Job Shop Scheduling). Still, to the best of our knowledge, we are not aware of a unified approach that can be easily modified to accommodate variations, while simultaneously being computationally tractable for large scale instances. On the contrary, it is widely believed that optimization might not be the right approach for certain classes of resource allocation problems such as scheduling for example. In fact, commercial solvers like ILOG for scheduling problems are typically not optimization based, but rather rule based.

Our aspiration in this paper is to develop a widely applicable, flexible and tractable modeling framework based on binary optimization that is capable of modeling and solving large scale instances for a variety of resource allocation problems over time. The paper has its intellectual origins with the work of Bertsimas and Stock (1998) on air traffic flow management, which is a resource allocation problem over time. In this problem, the resources are airports and sectors of the airspace, the requests are flights and the objective is to minimize delays in the system. To this date, within the scope of air traffic flow management, this modeling approach has proven successful, as it continues to be used extensively by several researchers and practitioners around the world. Two significant generalizations of the model in the context of air traffic flow management that suggest flexibility include: (a) the work of Bertsimas and Gupta (submitted for publication) that shows how fairness issues among airlines can be modeled in a computationally effective way, and (b) the work of Bertsimas, Lulli, and Odoni (2011) that allows the use of alternative routing of flights, when the current resources decrease possibly because of bad weather. Given the success of this modeling approach to air traffic flow management, it is natural to ask:

(a) Can we develop a modeling approach to general dynamic resource allocation problems that is flexible, tractable and widely applicable?
(b) Can we give some insights (both theoretical and empirical) on the reasons of the approach effectiveness in the context of resource allocation?

The broad framework we have in mind for *Dynamic Resource Allocation Problem* (DRAP) is as follows. The primitive quantities are: a set of resources $\mathcal{R}$ and a set of requests $\mathcal{I}$ belonging to a set of owners $\mathcal{O}$ that need to be processed by these resources over

* Corresponding author.
E-mail addresses: dbertsim@mit.edu (D. Bertsimas), shubhamg@mit.edu (S. Gupta), lulli@disco.unimib.it (G. Lulli).

0377-2217/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2013.10.063

a time horizon $\mathcal{T}$. Each request i needs to be completed by certain time and it can be completed by using alternative sets of resources $\mathcal{R}_i^1, \ldots, \mathcal{R}_i^{n_i}$. Different allocations of resources to requests over time result in delays if the request is completed after its desired time. The overall goal is to allocate resources to requests over time in order to complete the requests as efficiently as possible (minimum delay), potentially using alternative resources and ensuring that the distribution of delays amongst these requests (and implicitly to their owners) is fair.

In the opening sentence of his 1963 book Linear Programming and Extensions (Dantzig, 1963), George Dantzig writes: "the final test of any theory is its capacity to solve the problems which originated it". Motivated by this philosophy, we first present the modeling framework, and then use it to model and solve the following three widely-studied problems:

1. **Air Traffic Flow Management (ATFM).** Air Traffic Flow Management aims to prevent local demand-capacity imbalances by adjusting the flows of aircraft on a national or regional basis. ATFM models optimize for each flight the time of departure, the route selected, the time required to traverse each sector, and the time of arrival at the destination airport, taking into account the capacity of all the elements of the air traffic management system. In our generalization, each request will represent a scheduled sector-based path from takeoff through landing of a flight. The resources are the runways for takeoff and landing as well as the en-route airspace sectors. The fairness controls will be imposed at the destination airports where the final flight sequences are desired to be as close to the original scheduled sequence as possible. Finally, the alternative resources allows the option of flying a different origin–destination route for each aircraft.
2. **Aircraft Maintenance Problem (ACM).** The Aircraft Maintenance Problem is an important dynamic resource allocation problem in the airline industry. In this problem, an aircraft requests a large number of inspection and repair activities for which resources like equipments, sophisticated tools and highly specialized skills are used over several months.
3. **Job Shop Scheduling (JSP).** Job Shop Scheduling is one of the most notoriously hard combinatorial optimization problem. It entails processing a set of jobs on machines with the objective of minimizing some function of the completion times of the jobs (examples include the makespan, i.e., the maximum completion time and the minsum, i.e., the average completion time), subject to two requirements: (a) the sequence of machines for each job is prescribed; and (b) each machine can process at most one job at a time, and the schedule must be non-preemptive. In this paper, we model this problem within our framework and solve it for the two most widely applicable objective functions we mentioned earlier, makespan and minsum.

1.1. Literature review

In the literature, there is a lack of a scheduling framework of the form aspired in this research effort. Nonetheless, we attempt to enumerate the relevant papers which have the same taste. Bar-Noy, Bar-Yehuda, Freund, Naor, and Schieber (2001) present a unified approach to approximating resource allocation and scheduling. There has been recent work on using approximate dynamic programming (ADP) methods to solve dynamic resource allocation problems which overcome the "curses of dimensionality" of standard dynamic programming methods. Using ADP, Powell and Topaloglu (2005) describe solution strategies for large-scale resource allocation problems under uncertainty. Similarly, Gocgun and Ghate (2012) develop an ADP method that uses Lagrangian relaxation and constraint generation for dynamic stochastic resource scheduling problems. ADP techniques have also been used by Erdelyi and Topaloglu (2010) to solve a dynamic capacity allocation problem and by Powell and Van Roy (2004) who present computationally efficient algorithms for a mathematical model of dynamic resource allocation motivated by problems in transportation and logistics. In terms of specific applications, Menasce and Casalicchio (2004) design a framework for resource allocation in grid computing, whereas, Alhusaini, Prasanna, and Raghavendra (1999) focus on a unified resource scheduling framework for heterogeneous computing environments.

We next review relevant literature on the three problems discussed in this section.

- **ATFM.** This is an extensively studied problem. Starting with the first paper by Odoni (1987) in 1987, there have been a plethora of proposals attacking various aspects of the problem. One of the most comprehensive models is by Bertsimas and Stock (1998) which considers the problem of controlling release times and speed adjustments of aircraft while air-borne for a network of airports taking into account the capacitated airspace. For a detailed survey of the various contributions and a taxonomy of all the ATFM problems, see Bertsimas and Odoni (1997) and Hoffman, Mukherjee, and Vossen (2011).
- **ACM.** Please see Gharbi, Girard, Pellerin, and Villeneuve (1997) and Dijkstra, Kroon, Solomon, Van Nunen, and Van Wassenhove (1994) for details on this problem. For the special case of aircraft engine maintenance, Zarybnisky (2011) proposes near-optimal approximation algorithms. For F-series fighter aircraft (specifically, F-15 and F-100), Forbes and Wyatt (1975) and Amouzegar, Galway, and Geller (2002) contain details on the dependency between various maintenance components in the context of aircraft engine maintenance.
- **JSP.** There have been a plethora of proposals for JSP which utilize both exact methods and heuristic approaches. The earliest exact method can be traced back to Giffler and Thompson (1960) in 1960. Thereafter, many branch and bound type algorithms were developed by Carlier and Pinson (1989), Applegate and Cook (1991), Brucker, Jurisch, and Sievers (1994). An approach based on exploiting the disjunctive graph representation of JSP was developed by Adams, Balas, and Zawack (1988) and Balas and Vazacopoulos (1998). These are known as Shifting Bottleneck methods.

1.2. Structure of the paper

Section 2 describes the proposed binary optimization framework. It introduces the constraints for scheduling, usage of alternative resources and fairness controls. Section 3 enumerates several examples of problems and formulates the three applications mentioned in this section within our framework. Section 4 presents theoretical evidence (by providing several polyhedral insights) and Section 5 presents computational evidence on the strength of the overall framework. Section 6 contains concluding remarks.

1.3. Notation and preliminaries

We denote scalar quantities by lowercase, non-bold face symbols (e.g., $w \in \mathbb{R},\ k \in \mathbb{N}$), vector quantities by lowercase, boldface symbols (e.g., $\mathbf{w} \in \mathbb{R}^n,\ n > 1$), and matrices by uppercase, boldface symbols (e.g., $\mathbf{A} \in \mathbb{R}^{n \times n},\ n > 1$).