A Lagrangian Heuristic for Integrated Timetabling and Vehicle Scheduling

Antonio Frangioni · Laura Galli · Samuela Carosi · Leopoldo Girardi · Giuliano Vallese

Abstract In this paper we propose a Lagrangian heuristic approach to solve the *integrated timetabling and vehicle scheduling* problem. The approach was developed in a tight collaboration between Università di Pisa and M.A.I.O.R, a company producing software for public transport authorities and operators, which also provided real-word test instances.

Keywords Timetabling · Vehicle Scheduling · Integration · Math-heuristic

1 Introduction

Planning of a public transportation system is a complex process that consists of several phases, such as strategic planning (e.g., network design), tactical planning (e.g., line planning, timetabling), operational planning (e.g., vehicle/crew scheduling) and real time control. Due to the hardness of the problem, these phases are usually performed in sequence, in order to reduce its complexity. In this way the number of vehicles and drivers needed, expensive resources that need efficient utilization, is determined last. Despite numerous studies addressing each step individually, to our knowledge, the literature for *integrated* approaches is rather scarce, e.g., [1, 2, 4, 3]. In this paper we propose a Math-heuristic approach to solve the *integrated timetabling and vehicle scheduling* problem, with a special focus on buses. The approach was developed in a tight collaboration between Università di Pisa and M.A.I.O.R, a company producing software for public transport authorities and operators, which also provided real-word test instances.

The system is already fully implemented and functional, and is part of the *MAIOR* commercial suite.

Antonio Frangioni, Laura Galli
Dipartimento di Informatica, Università di Pisa E-mail: frangio@di.unipi.it

Samuela Carosi, Leopoldo Girardi, Giuliano Vallese
M.A.I.O.R. Srl E-mail: samuela.carosi@maior.it, giuliano.vallese@maior.it

Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling (PATAT-2018), Vienna, Austria, August 28–31, 2018

2 Problem description

A *public transportation network* PTN is given, usually under the form of a graph where the nodes correspond to stops or depots, and the links correspond to direct bus transits. A *line* l is a path in the PTN between two terminus A and B; generally, lines run in both directions, i.e., a line is actually a pair of lines. The *frequency* of a line specifies how often bus service should be offered, as measured by the *ideal time interval* between two subsequent transits at a specific point, called "pilot node". The time horizon (one day) is *aperiodic* and is divided into time slots, in which the service frequency typically varies.

We assume that the set of lines L is already given, together with the desired frequencies. *Trips* are paths, corresponding to lines in the PTN, which have to be operated at a certain time. Each trip $c \in C$ is characterized by a start and end time, as well as the arrival time π_c at the pilot node. The problem takes as input a set of trips $C = \cup_{l \in L} C_l$, partitioned according to the different lines ($\cap_{l \in L} C_l = \emptyset$), and it can be divided in two parts. The *timetabling (TT)* problem consists of selecting, for each line l, a subset of trips $\bar{C}_l \subset C_l$, such that the corresponding frequencies are as close as possible to the desired ones. The goal of *bus scheduling (BS)* is instead to find a cost-minimal assignment between buses and trips such that each trip is covered by exactly one bus and the schedule is "feasible"; this entails *time compatibility* between ending and starting time of consecutive trips, and the possible impact of regulations.

Overall, this results in the *bus routes* together with a *timetable* for each of the lines. The goal of the integrated problem is to provide a solution that optimally balances *cost* (i.e., minimize the number of buses used) and *user satisfaction* (i.e., minimize the distance between actual and desired frequencies).

3 Solution approach

The *TT problem* is separable for each $l \in L$, as frequencies are defined "per line". From the trips C_l we construct a directed graph $G_l^{TT} = (N_l^{TT}, A_l^{TT})$, where the nodes correspond to the trips, plus two nodes O^+ and O^- representing the start and the end of the day for the line. Then, we define three types of arcs: (i) (c_1, c_2), meaning that the two trips can be subsequent trips of the line; (ii) (O^+, c), meaning that the c trip can be the first of the day, (iii) (c, O^-), meaning that the c trip can be the last of the day. The cost of a type (i) arc depends on the difference between the ideal time interval (in the corresponding time slot) and $\pi_{c2} - \pi_{c1}$, with a simple formula whose details are not crucial. The arc exists only if $\pi_{c2} - \pi_{c1}$ belongs to a given interval, i.e., pairs of trips "too close" or "too far apart" cannot be chosen. The cost of a type (ii) and (iii) arc depends on the difference between the starting (respectively, ending) time of the day and π_c. The arc exists only if π_c is "close enough" to the starting (ending) time of the day for the line.

It is trivial to see that G_l^{TT} is acyclic; therefore, the TT problem can be easily solved, for each line l, as an *acyclic shortest path (SP)* problem.

Proceedings of the 12th International Conference on the Practice and Theory of Automated Timetabling (PATAT-2018), Vienna, Austria, August 28–31, 2018

The *BS problem* is not separable, as a single bus route can cover different lines. We construct a single *compatibility graph* $G^{BS} = (N^{BS}, A^{BS})$ having two nodes c^- and c^+ for each trip $c \in C$, that represent the start and the end of trip c, respectively, plus two nodes O^+ and O^- representing buses leaving and returning to the depot, respectively. Then, we define four types of arcs: (i) (c^-, c^+), meaning that the corresponding trip c is covered by a bus; (ii) (c^+, d^-), the *compatibility* arc defined iff the same bus route can cover trip c and then trip d in sequence; (iii) (d^+, O^-), meaning that the bus returns to the depot right after completing trip d, and (iv) (O^+, c^-), meaning that the bus starts trip c right after leaving the depot (and reaching the corresponding terminus). Finally, a return arc (O^-, O^+) is added to define a circulation problem, whose capacity is equal to the fleet cardinality and whose cost represent the cost of a bus leaving the depot. The cost of a type (ii) arc depends on the difference between the minimum recovery time (in the corresponding time slot) and $t_{d^-} - t_{c^+}$, with a simple formula whose details are not crucial. The arc exists only if $t_{d^-} - t_{c^+}$ belongs to a given interval, i.e., pairs of trips "too close" or "too far apart" cannot be chosen.

Clearly, the BS problem can be solved as a *minimum cost network flow (MCF)* on G^{BS}, but taken alone its optimal solution would be the all-0 flow since there is no constraint requiring trips to be covered.

All in all, the integrated model combines the BS graph and the TT graphs (one for each line) to yield the following Mixed Integer Linear Programming (MILP) model:

$$\min \ \alpha x + \textstyle\sum_{l \in L} \beta^l y^l \qquad (1)$$

$$\textstyle\sum_{(j,i) \in A^{BS}} x_{ji} - \sum_{(i,j) \in A^{BS}} x_{ij} = 0 \qquad i \in N^{BS} \qquad (2)$$

$$\textstyle\sum_{(j,i) \in A_l^{TT}} y_{ji}^l - \sum_{(i,j) \in A_l^{TT}} y_{ij}^l = b_i^l \qquad l \in L,\ i \in N_l^{TT} \qquad (3)$$

$$0 \le x_{ij} \le u_{ij} \qquad (i,j) \in A^{BS} \qquad (4)$$

$$y_{ij}^l \in \{0,1\} \qquad l \in L,\ (i,j) \in A_l^{TT} \qquad (5)$$

$$\textstyle\sum_{(i,j) \in B(c)} y_{ij}^l = x_{c^-,c^+} \qquad l \in L, c \in C_l \qquad (6)$$

In the formulation, (2) and (4) represent the MCF problem, where the capacities u_{ij} for BS are all one except that of the return arc (O^-, O^+); (3) and (5) represent the SP problems, where the deficits b_i^l for TT are all zero except in the source/sink nodes; (6) are the linking constraints between BS and TTs, where for each $l \in L, c \in C_l$, the set $B(c)$ contains all arcs of A_l^{TT} entering the node representing c.

Choosing the coefficients α and β of the objective function is nontrivial, as correctly doing so is crucial for obtaining a compromise between the two contrasting objective functions of the problem.

Our solution approach relies on the Lagrangian relaxation of the linking constraints (6): relaxing (6) we obtain one MCF and $|L|$ SP independent problems that can be efficiently solved. The Lagrangian dual is solved using a *Bun-*

dle method. The algorithm uses the primal information of the relaxations to guide a fixing heuristic that progressively constructs integer solutions, i.e., the bus routes.

The idea behind the heuristic is to build the solution in an orderly fashion, starting from the beginning of the day and moving towards the end of it. This is done by studying the forward stars (FS) in the TT subproblems of the nodes associated to sources or already fixed trips. Using the value given to each arc by the continuous solution of the Lagrangian relaxation, we can determine which trip is more suitable to be the first of the day for a line (if we are studying the FS of a source node) or the next one (if we are studying the FS of an already fixed node).

4 Computational Results

We tested our algorithm on 12 real-world instances provided by MAIOR, with different number of lines, using an Intel (R) Xeon (R) CPU E5-2420 1.9Ghz, see Table 1. We consider two versions of our heuristic: "h-*Bundle*" solves the Lagrangian dual using the *Bundle* method, while "h-`Clp`" solves the continuous relaxation using the `Clp`, an open-source LP solver. We call "BSol" (i.e., best solver), the heuristic that produces the best solution, and "BTS" (i.e., best time solver) the corresponding time (ranging from a few minutes to 6 hours). We compare the solution produced by BSol with solutions produced manually by timetabling experts ("*Manual*"), as well as using the `CPLEX` MILP solver (V12.7) on the full formulation (1)-(6) with time limit corresponding to 1, 2 and 4 times BTS. The results show the *percentage gain* "$X\%$" of the objective value of our solutions. On small instances h-`Clp` performs better, while on larger instances h-*Bundle* is often preferable. The obtained solutions are much better than the manually obtained ones, and the improved quality is perceived by experts of the field. `CPLEX` sometimes finds better solutions ("$-X\%$" highlighted in bold), but in general the heuristic approach is competitive. Note that "$X+$" means that the cost of the solution is X times larger, and "nA" means that CPLEX could not find any feasible solution within the time limit. We can conclude that our method is competitive on real-world instances with respect to both manual solutions and a general MILP solver like `CPLEX`.

References

1. Desaulniers, G., Hickman, M.: Public Transit. In: C.B. G. Laporte (ed.) Transportation, Handbooks in Operations Research and Management Science, 2006 (2006)
2. Michaelis, M., Schöbel, A.: Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. Public Transport **1**(1), 211–232 (2009)
3. Sun, D., Xu, Y., Peng, Z.R.: Timetable optimization for single bus line based on hybrid vehicle size model. Journal of Traffic and Transportation Engineering **2**(3), 179–186 (2013)
4. Wagale, M., Singh, A.P., Sarkar, A.K., Arkatkar, S.: Real-time Optimal Bus Scheduling for a City Using A DTR Model. Procedia - Social and Behavioral Sciences **104**(1), 845–854 (2013)

Table 1 Comparison of solution quality: heuristics, manual and `CPLEX`.

#Lines	Instance	BSol	BTS(s)	O.V.				
				Manual vs h-Clp	h-*Bundle* vs h-Clp	Cplex vs BTS*1	Cplex vs BTS*2	Cplex vs BTS*4
2	2A	h-Clp	100	19.04%	-4.64%	nA	**-7.88%**	**-7.88%**
	2B	h-Clp	477	35.35%	-0.93%	14+	14+	0.71%
	2C	h-Clp	191	51.78%	-0.41%	11+	8+	25.72%
	2D	h-Clp	335	117.93%	-3.53%	nA	nA	4.60%
	2E	h-Clp	611	202.99%	-2.31%	30+	2.06%	**-1.50%**
	2F	h-Clp	804	28.84%	-5.81%	18+	0.75%	0.75%
	2G	h-Clp	762	158.86%	-30.17%	53+	9+	**-16.23%**
4	4A	h-Clp	2837	81.38%	-3.58%	nA	19.61%	19.61%
	4B	h-*Bundle*	3156	16.86%	0.46%	nA	nA	6.11%
	4C	h-Clp	6947	252.98%	-4.39%	nA	48.84%	48.61%
	4D	h-*Bundle*	11725	80.86%	10.60%	nA	nA	nA
8	8A	h-*Bundle*	26250	68.60%	4.37%	nA	nA	nA