Integrating Crew Scheduling and Crew Rostering for rail freight with train delays

Héloïse Gachet, Frédéric Meunier, Juliette Pouzet

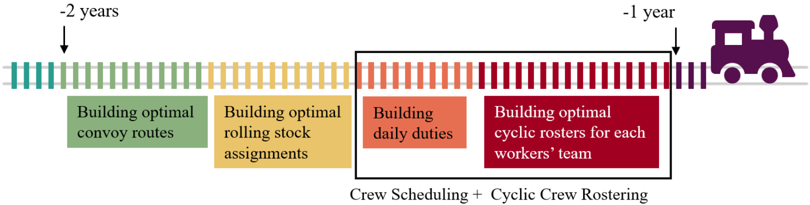
November 18th, 2025

École nationale des ponts et chaussées

Rail freight in France

- Modal share for rail freight: 18% in Europe (10% in France)
- Between 1800 and 2000 trains per week
- Many differences with passengers transportation:
  - Priority goes to passengers
  - Trains mostly at night

Resource planning



Contributions

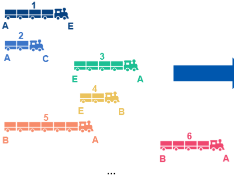
We propose a complete methodology:

- modeling the "resilience" of rosters to train delays at the SNCF
- solving the integrated problem while handling train delays without sampling
- providing a tight lower bound
- leading to an average 8% cost reduction for the deterministic version

Problem definition

Input: Trains on a typical week

Output: Covering of trains by "blocks" with minimum cost, each block assigned to a team (each block then placed in a cyclic roster)



Team A:

sun	mon	tue
duty 7	duty 1	on-call
10h-17h	12h-21h	
A - A	A - A	

sat	sun	mon	tue
duty 5	duty 6	duty 2	duty 3
11h-17h	5h-16h	12h-19h	5h-11h
A - E	E - A	A - B	B - A

Team B:

mon	tue	wed
duty 4	duty 8	duty 9
12h-20h	11h-17h	5h-12h
B - B	B - B	E - B

cost = 10

Team A:

mon	tue	wed	thu	fri	sat	sun
duty 1	on-call	R	R	R	duty 5	duty 6
12h-21h					11h-17h	5h-16h
A - A					A - E	E - A
duty 2	duty 3	R	R	R	R	duty 7
12h-19h	5h-11h					10h-17h
A - B	B - A					A - A

Team B:

mon	tue	wed	thu	fri	sat	sun
duty 4	duty 8	duty 9	R	R	R	R
12h-20h	11h-17h	5h-12h				
B - B	B - B	E - B				

Stochastic train delays: context

Train departure and arrival delays (primary delays and secondary delays) $\rightarrow$ disruptions of rosters with propagation within duties.

When delays exceed some threshold, disruptions occur and they induce prohibitive adjustments:

- a full duty, or some of its trains, must be operated by backup drivers;
- an on-call must be added within a block or at its end.

Stochastic train delays: fragility and assumptions

Assumptions:

- primary delays are independent variables (not necessarily identically distributed),
- their distributions are known,
- their support is finite.

H: random variable for arrival times (with primary and secondary delays)
$\sigma(h, b)$: number of disruptions in the realization $H = h$.

Roster "fragility" score:

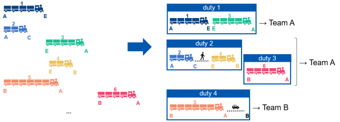
$$\text{fragility}(r) := \mathbb{E}\Big[\sum_{b \in r} \sigma(H, b)\Big]$$

Heuristic: sequential approach

❶ Crew Scheduling

Input: Trains on a typical week

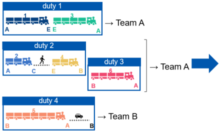
Output: Covering of trains by daily duties with minimum cost, each duty assigned to a team



❷ Crew Rostering

Input: Duties, each assigned to a team

Output: Covering of trains by blocks with minimum cost, each block assigned to a team (then placed in a roster)



Team A:

sat	sun	mon	tue
on-call	duty 7 5h-14h A - A	duty 1 12h-21h A - A	on-call

sat	sun	mon	tue
duty 5	duty 6	duty 2	duty 3
11h-17h	11h-18h	12h-19h	5h-11h
A - E	E - A	A - B	B - A

Team B:

mon	tue	wed
duty 4	duty 8	duty 9
12h-20h	11h-17h	5h-12h
B - B	B - B	E - B

cost = 11

→ Leads to sub-optimal solutions

→ Hard to account optimally for uncertainties

Model

$$\begin{aligned} \text{minimize} \quad & \sum_{i\in I}\sum_{b\in B_i} c_b x_{b,i} \\ \text{s.t.} \quad & \sum_{i\in I}\sum_{b\in B_i : t\in b} x_{b,i} \geqslant 1 & \forall t \in T \\ & \sum_{i\in I}\sum_{b\in B_i} \mathbb{E}[\sigma(H,b)] x_{b,i} \leqslant S \\ & x_{b,i} \in \{0,1\} & \forall i \in I, \forall b \in B_i \end{aligned}$$

B_i = set of feasible "blocks" for the team i

$\rightarrow$ combinatorial explosion

National input:

1800 trains $\rightarrow \sim 10^{25}$ blocks

Column Generation: standard resolution methodology of a linear program when the number of variables is large

Column Generation

$$\begin{aligned}
\text{minimize} \quad & \sum_{i\in I}\sum_{b\in\overline{B}_i} c_b x_{b,i} \\
\text{s.t.} \quad & \sum_{i\in I}\sum_{b\in\overline{B}_i : t\in b} x_{b,i} \geqslant 1 \qquad \forall t\in T \\
& \sum_{i\in I}\sum_{b\in\overline{B}_i} \mathbb{E}[\sigma(H,b)] x_{b,i} \leqslant S \\
& x_{b,i}\in\{0,1\} \qquad \forall i\in I, \forall b\in\overline{B}_i\subset B_i
\end{aligned}$$

Solving the linear relaxation with B_i:

- Iteratively
  - solving the linear relaxation with $\overline{B}_i$
  - retrieving dual information
  - searching for solutions to a "pricing sub-problem"
  - adding elements to $\overline{B}_i$
- Optimality guarantee when no elements can be added

Solving the integer problem with B_i:

- Upper bound and feasible solution with $\overline{B}_i \subseteq B_i$

Challenge: Solving the pricing sub-problem quickly

For our problem:

Pricing sub-problem = Shortest path with constraints (Resource Constrained Shortest Path)

Resource Constrained Shortest Path: setting

Input: Graph with vertices o and d, feasible $o-d$ paths $\mathcal{P}_{od}$ and a (non-linear) cost function c
Ouput: Feasible $o-d$ path $P \in \mathcal{P}_{od}$ with minimum cost $c(P)$

Resources: Vector tracking accumulated valuable quantities along a path through an extension (non-linear) function

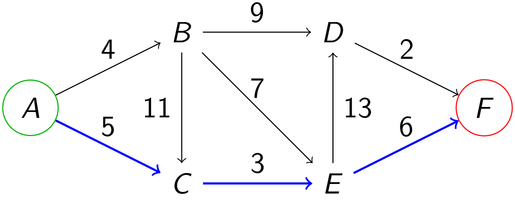
e.g.: shortest $o-d$ path under time budget
$\rightarrow$ *resources: accumulated length, accumulated arc time*
e.g.: shortest $o-d$ path under stochastic time budget
$\rightarrow$ *resources: accumulated length, accumulated arc time distribution*

Shortest path

Usual shortest path

Input: Graph with vertices o and d, non-negative cost on each arc

Ouput: $o - d$ path with minimum cost



$\rightarrow$ Dijkstra's algorithm

$\rightarrow$ A^* search algorithm

Shortest path

A^* search algorithm

Principle: Enumeration algorithm with bounds to discard paths

- Bound b_v under-estimating cost of shortest path from any vertex v to d
- Discard paths P from o to v with "estimated cost" $c(P) + b_v$ greater than one of an explored $o - d$ path

$\rightarrow$ with $b_v = 0$: Dijkstra's algorithm

e.g.: shortest route on a map
$\rightarrow b_v =$ *distance as the crow flies*

Resource Constrained Shortest Path: algorithms

Resolution: Enumeration algorithm using

- Key: order of paths processing
- Bound: under-estimate of the resources and cost to reach d
- [not in A^*] Dominance: comparison of resources for paths with same cost

Different algorithms:[1]

- Generalized A^*

  Key = "estimated cost", discard paths using "estimated cost"
- Label dominance

  Key = cost of path, discard paths using dominance
- Label correcting

  Key = "estimated cost", discard paths using "estimated cost" and dominance

[1] synthetized by A.Parmentier in Algorithms for Non-Linear and Stochastic Resource Constrained Shortest Paths, 2017.

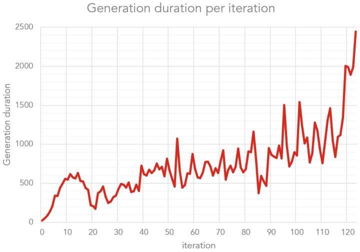
For our problem

How to handle the duty and block feasibility constraints and the stochastic constraint?

- One graph per team, with trains as vertices and arcs encoding possible successions of trains.
- Resources keep track of several indicators, as:
  - block indicators (number of days, etc.)
  - duties indicators (range, driving duration, etc.)
  - delay distributions

Results: without train delays

	Sequential approach		Our approach		
	Objective	Time	Obj. lower bound	Obj. upper bound[2]	Total time
Instance 280 trains	228	3min	205.3	**209 (-8%)**	21h07min
Instance 925 trains	832	26min	711.9	**730 (-12%)**	23h48min
Instance 1810 trains	1633	1h04min	1395.0	**1494 (-8%)**	40h16min



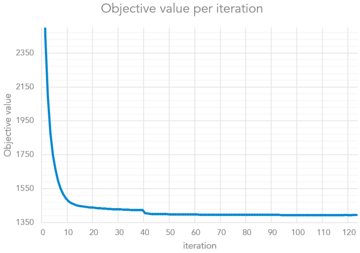


Figure: Column Generation indicators per iteration

[2] solving MILP with CG output variables and a 1h time limit

Next goals

- Gap closing (or strenghtening) for large instances without train delays
- Getting results with train delays on small and large instances
- Challenging the finite support assumption on the distributions

Thank you!