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

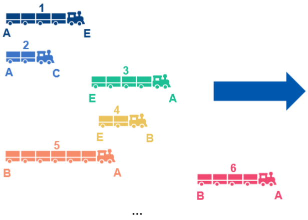
Héloïse Gachet, Frédéric Meunier, Juliette Pouzet

February 24th, 2026

Problem definition: deterministic

Input: Trains on a typical week, teams of workers

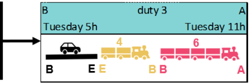
Output: Covering of trains by "blocks" with minimum number of working days, each block assigned to a team (each block then placed in a roster)



Team A-1:

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-1:

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

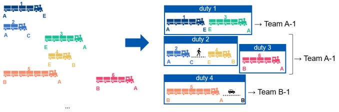
number of working days = 10

Current heuristic: two-stage approach

❶ Crew Scheduling[1]

Input: Trains on a typical week, teams of workers

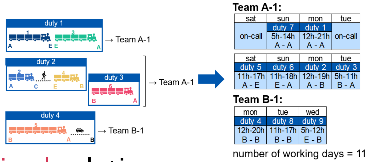
Output: Covering of trains by daily duties with minimum surrogate cost, each duty assigned to a team



❷ Crew Rostering[2]

Input: Duties, each assigned to a team

Output: Covering of trains by blocks with minimum number of working days, each block assigned to a team



→ Leads to sub-optimal solutions

→ Hard to account for uncertainties

[1] Heil, J., Hoffmann, K., Buscher, U.: Railway crew scheduling: Models, methods and applications. (2020)

[2] Kohl, N., Karisch, S.E.: Airline Crew Rostering: Problem Types, Modeling, and Optimization. (2004)

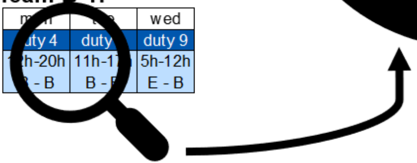
Behind the word “uncertainties”

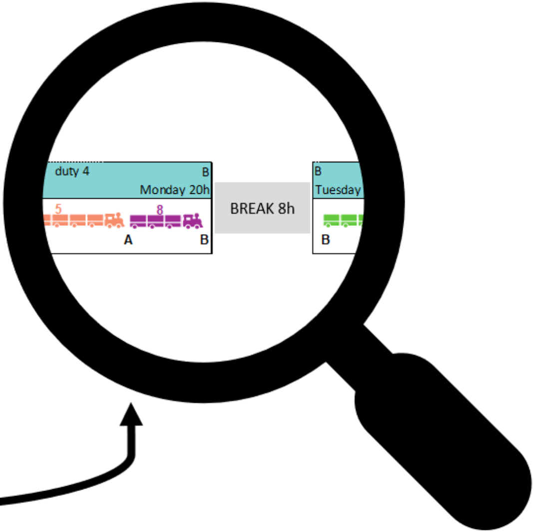
Team A-1:

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-1:





Complete problem definition

Input: Trains on a typical week, teams of workers, threshold S
Output: Covering of trains by "blocks" with minimum number of working days, each block assigned to a team, such that the overall "fragility" is under the threshold S

"Fragility": metric we propose to evaluate the impact of train delays on blocks (formally introduced later)

- Surveys on the problem without uncertainties in airline and railway industries[1]
- Robust or stochastic optimization for the Crew Scheduling in airline industries[2]

[1] Mertens, L., Wolbeck, L.A., Rößler, D., Xie, L., Kliewer, N.: An overview of optimization approaches for scheduling and rostering resources in public transportation. (2023)

[2] Dück, V., Ionescu, L., Kliewer, N., Suhl, L.: Increasing stability of crew and aircraft schedules. (2012)

Contributions

- Introduction of the “fragility” metric measuring the potential impact of train delays on blocks
- Complete methodology solving the deterministic problem and the complete problem (with train delays)
  Can be turned into an exact method when either
  - delay distributions have finite support
  - an oracle assumption holds (and delay distribution are arbitrary)
- Substantial improvements over the current solution at the SNCF, for both the deterministic problem and the complete problem (with train delays)

Modeling

- Deterministic problem as an integer linear program, to be solved by column generation
- "Fragility" of blocks to train delays
- Complete problem as an integer linear program, with an extra "fragility" stochastic constraint, to be solved by column generation

Model: deterministic problem

- w_b: number of working days in block b
- I: set of teams
- B_i: set of all feasible blocks for team i
- T: collection of trains

$$\begin{aligned} \text{minimize} \quad & \sum_{i \in I} \sum_{b \in B_i} w_b x_{b,i} \\ \text{s.t.} \quad & \sum_{i \in I} \sum_{b \in B_i : t \in b} x_{b,i} \geqslant 1 \quad && \forall t \in T \\ & x_{b,i} \in \{0,1\} && \forall i \in I, \forall b \in B_i \end{aligned}$$

- $x_{b,i} \in \{0,1\}$: block $b \in B_i$ is selected for team $i \in I$

Uncertainties: train delays

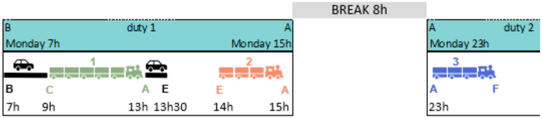
Train delays of two kinds:

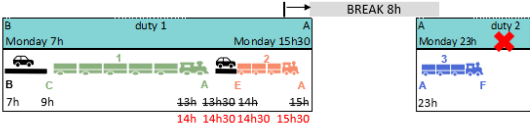
- Primary delays: exogenous
- Secondary delays: endogenous, appear from delay propagation within duties

When delays exceed some threshold:

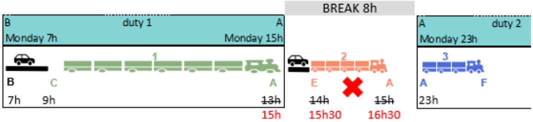
- a break might not meet the minimal duration. In this case, the next duty is disrupted;
- the working time within the duty might be too long. In this case, one or multiple trains are disrupted.

Train delays: example





duty 2 is disrupted



train 2 is disrupted in duty 1

Fragility score

Assumption:

- primary delays distributions are known

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

Block “fragility” score:

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

Model: complete problem

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

- $x_{b,i} \in \{0, 1\}$: block $b \in B_i$ is selected for team $i \in I$

Methodology

- Column generation for both the deterministic and the complete problems
  Assumptions:
  - primary delays are independent variables (not necessarily identically distributed),
  - their supports are finite.
- A method to address any delay distribution in an exact way (under an extra oracle assumption)

Column generation

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

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 Problem (RCSPP))

RCSPP: setting

Input:

- Acyclic digraph with vertices o and d, partial ordered set of resources $(\mathcal{R}, \preccurlyeq)$
- Non-decreasing "extension function" on each arc $f_a \colon \mathcal{R} \to \mathcal{R}$ (describes evolution of a resource when arc a is added to a path $r_{P+a} := f_a(r_P)$) NON LINEAR
- Non-decreasing "cost and feasibility" function $c \colon \mathcal{R} \to \mathbb{R} \cup \{+\infty\}$ NON LINEAR

Output: Feasible o-d path P with minimum cost $c(r_P)$

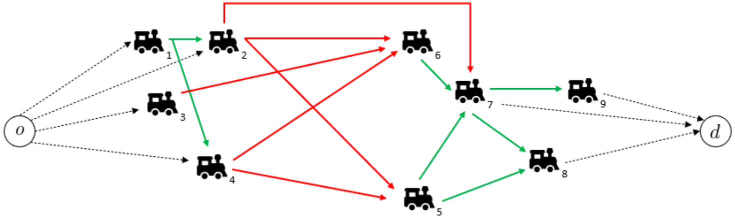
e.g., shortest o-d path under time budget

$\rightarrow$ *resource: (total arc cost, total arc time)*

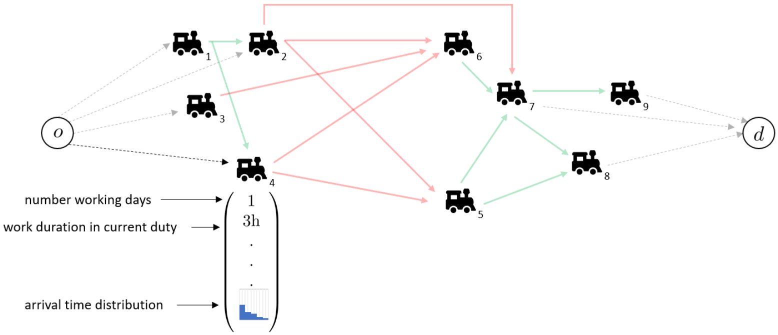
$\rightarrow$ *extension function: (+arc cost, +arc time)*

$\rightarrow$ *cost and feasibility function: total arc cost if total arc time $\leqslant$ budget, $+\infty$ else*

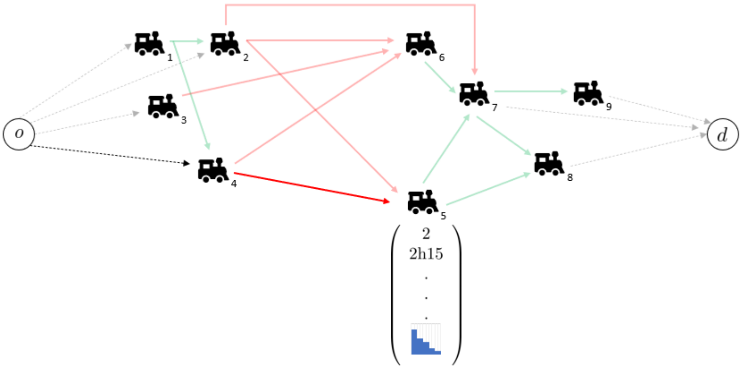
RCSPP: for the complete version



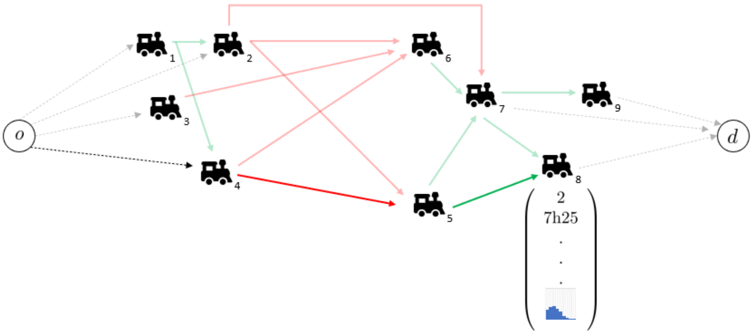
RCSPP: for the complete version



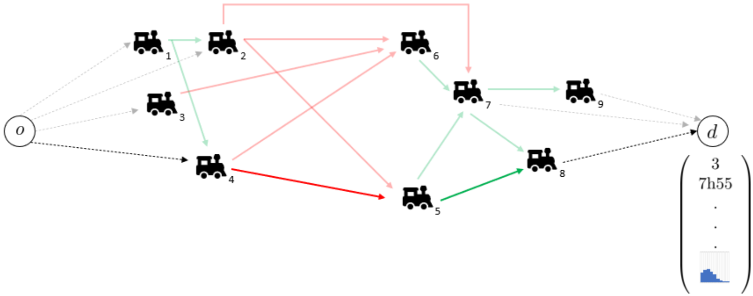
RCSPP: for the complete version



RCSPP: for the complete version



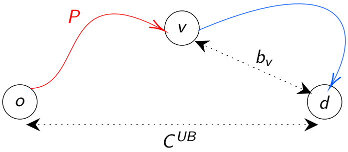
RCSPP: for the complete version



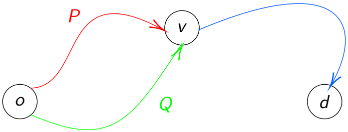
RCSPP: algorithms

Resolution:[3] Enumeration algorithm using

- Bounds: under-estimate of the resource to reach d

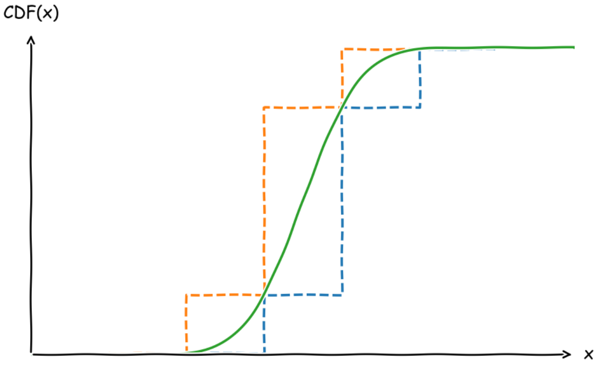


- Dominance: comparison of resources for paths with same last vertex



[3] synthetized by Parmentier, A.: Algorithms for Non-Linear and Stochastic Resource Constrained Shortest Paths. (2017)

Addressing any delay distribution



$$P'_t \leqslant_{\mathrm{st}} P_t \leqslant_{\mathrm{st}} P''_t$$

optimal value under $(P'_t)_t \leqslant$ optimal value under $(P_t)_t \leqslant$ optimal value under $(P''_t)_t$

Addressing any delay distribution

- Original primary delay distributions $(\bar{P}_t)_t$
- Sequence of alternative primary delay distributions $(P_{t,n})_{t,n}$ such that
  - $P_{t,n} \leqslant_{\mathrm{st}} \bar{P}_t$ for every t and n
  - $P_{t,n} \xrightarrow{d} \bar{P}_t$ for every t

Theorem G.–Meunier 2025+

Suppose that the problem is feasible under $(\bar{P}_t)_t$. Then there exists N such that, for all $n \geqslant N$, the sets of feasible solutions under $(P_{t,n})_t$ and under $(\bar{P}_t)_t$ coincide (and the same holds for the set of optimal solutions).

Results: deterministic problem

- Run with:
  - Intel(R) Xeon(R) CPU E5-2690 v2 @ 3.00GHz processor
  - programming language Java 18
  - solver CPLEX 22.11
- Column generation time limit 48 h
- Integer linear program time limit 1 h

primary delay value	probability	
0 min	0.5	P^A
15 min	0.2	P^A
30 min	0.2	P^A
60 min	0.1	P^A
0 min	0.2	P^B
15 min	0.1	P^B
30 min	0.2	P^B
45 min	0.2	P^B
60 min	0.2	P^B
90 min	0.1	P^B

	Two-stage approach			Integrated approach: deterministic case			
	Time	Objective	Fragility	Time	LB	UB	Fragility
116 trains (P^A)	1 min	50	0.970	15 h 13 min	40.3	41	3.116
166 trains (P^A)	1 min	102	1.950	12 h 31 min	90.5	92	3.633
280 trains (P^A)	3 min	228	2.450	19 h 10 min	205.3	207	5.829
926 trains (P^A)	26 min	832	8.538	30 h 50 min	711.9	773	34.565
1,815 trains (P^A)	64 min	1633	14.900	45 h 30 min	1,395.0	1,548	46.305
1,815 trains ($P^A + P^B$)	64 min	1633	18.410	45 h 30 min	1,395.0	1,548	58.995

Results: complete problem (with train delays)

- Run with:
  - Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz processor
  - programming language Java 18
  - solver CPLEX 22.11
- Column generation time limit 96 h
- Integer linear program time limit 1 h

primary delay value	probability	
0 min	0.5	P^A
15 min	0.2	
30 min	0.2	
60 min	0.1	
0 min	0.2	P^B
15 min	0.1	
30 min	0.2	
45 min	0.2	
60 min	0.2	
90 min	0.1	

	Two-stage approach			Integrated approach: with uncertainties				
	Time	Objective	Fragility	Threshold S	Time	Best RMP	UB	Fragility
116 trains (P^A)	1 min	50	0.970	0.485	/	45.0	47	0.241
166 trains (P^A)	1 min	102	1.950	0.950	/	95.3	98	0.863
280 trains (P^A)	3 min	228	2.450	1.225	/	214.5	219	1.090
926 trains (P^A)	26 min	832	8.538	4.269	/	742.1	801	3.203
1,815 trains (P^A)	64 min	1633	14.900	7.450	/	1,451.2	1,592	7.321
1,815 trains ($P^A + P^B$)	64 min	1633	18.410	9.205	/	1,501.4	1,629	9.061

Next goals

- Close the gap for the deterministic version
- Lower bound for the complete version (with train delays)

Open Question

Is feasibility decidable for the complete problem under arbitrary delay distributions?

Thank you!