



On the Right Track Optimisation Models for Railway Planning

Janny M.Y. Leung,^{*} Yong-Hong Kuo[†]
and David S.W. Lai[‡]

Tuen mun subway
© Huating | Dreamstime.com

There are one million kilometres of railways in the world servicing over three trillion (10^{12}) passenger-kilometres per year. From urban public transit systems to suburban/regional railways to high-speed intercity trains, millions of people rely on railways for work, leisure and travel. Therefore, making sure that the trains get the passengers to their destinations safely, affordably, comfortably and on time is not an easy task.

What makes up a railway system? What we can see is the infrastructure (the so-called fixed stock of the network of railway tracks, the stations, the yards, the signalling systems, etc.) and the rolling stock of locomotives, carriages and wagons that move on the railways carrying passengers and goods. From the steam locomotives of the 19th century to the magnetic-levitation trains of today, technological advances have substantially increased the speed and energy efficiency of trains. Just as important as engineering improvements to the railway equipment, advances in the planning of the operations of the railway system have increased the capacity, comfort and convenience of railway transport.

The operational plan for a railway system is too complex to be solved as a monolithic problem, and is typically tackled in stages. The key stages are:

1. Network planning – determining the location and connectivities of tracks, stations and yards.
2. Line planning – determining the origin/destination termini and route of lines and their frequencies so as to provide sufficient capacity for the passenger trips to be served.

3. Timetabling – determining the despatch times of trains for each line; ensuring minimal waiting times and transit times for *all* passengers is particularly difficult.
4. Vehicle scheduling – assignment of locomotives and coaches to trips, and
5. Crew scheduling – rosters and duties for the drivers are determined, work rules and skills compatibilities make the assignment of operators to trips a difficult problem.

In this article, we will discuss how complicated some of these planning problems are and how the use of mathematical modelling and analysis enables these very large-scale complex problems to be solved. We will also discuss some recent trends in railway planning and operations.

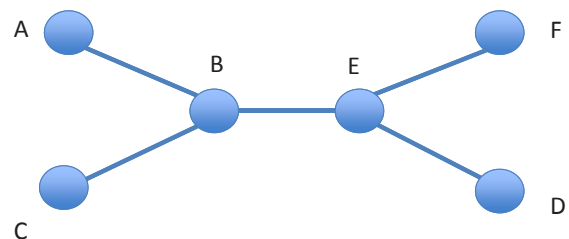


Figure 1

^{*}School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen)

[†]Stanley Ho Big Data Decision Analytics Research Centre, The Chinese University of Hong Kong

[‡]Department of Information, Logistics and Innovation, Vrije Universiteit Amsterdam

1 Network and line planning

Few cities have the luxury to build an entirely new railway system, so the infrastructure is often grown incrementally, subject to economic, environmental and political restrictions. Having to work with the existing rail network, one way to accommodate the changing passenger travel demand as a city grows and activity patterns change is to reorganise the termini and routes of the trains. This problem is called line planning.

Nowadays, with the prevalence of smart card ticketing systems (e.g. Octopus card in Hong Kong, Oyster card in London, the OV Chipkaart in the Netherlands), accurate estimation of the origin-destination (O-D) demand pattern for passengers by day-of-week, even by time-of-day, can be obtained. Consider a simple network as shown in Figure 1. With passengers travelling between every pair of vertices. All the passenger trips can be served with two lines A–B–E–D and C–B–E–F, although some passengers might have to change trains. Given a railway network, each possible path in the network is a possible line. Generally, there are too many paths to consider, so only a subset of such paths are candidates for lines. The line-planning problem is to choose the frequency x_l for each line l . Due to safety concerns, there is an upper bound on the frequency, and service considerations would impose a lower bound on x_l . The capacity of the line is given by $\gamma_l x_l$ where γ_l is the capacity of the train used on the line (depends on the number and type of coaches). The set of lines and frequencies should be chosen so the anticipated passenger demand is covered.

Generally, passengers would prefer not to have to change trains in their trips, so a possible objective for line planning is to maximise the number of direct trips for all passengers. This leads to a mixed-integer programming problem of large size, with thousands of constraints and hundreds of thousands of variables.

... Most urban transit rail networks are complex ... Balancing the arrival/departure times of trains at all the interchange stations to try to minimise the waiting times for all interchanging passengers is a very difficult task ...

assignment of time $\pi : V \rightarrow [0, T)$ that satisfies:

$$(\pi_j - \pi_i - \ell_a) \bmod T \leq u_a - \ell_a, \quad \forall a = (i, j) \in A.$$

These constraints can be used to model key aspects of railway operations, such as the sequence of arrivals/departures at stations of each line, matching timings for interchanges between lines, safety distances between trains, and sharing of tracks among lines. PESP can also be reformulated as a mixed-integer programming problem. Using this type of model, researchers have helped to design the timetable for the public transit system in Berlin [2].

For high-frequency inner-city metro lines, aperiodic timetables are sometimes used and can be tuned to offer higher service levels than periodic timetables. Most urban transit rail networks are complex and passengers may need to make several interchanges between lines to reach their destination. Balancing the arrival/departure times of trains at all the interchange stations to try to minimise the waiting times for *all* interchanging passengers is a very difficult task. Holding back one train at a station to

accommodate connecting passengers may lead to other passengers missing their connections further down the line.

Again, mixed-integer programming models can be used to address this problem, although additional binary variables and constraints have to be introduced to be able to correctly represent the waiting times for the next available train. Because the timetable

need not be periodic, the matching up of trains at interchange stations has to be determined for all *possible* connections, leading to a substantial increase in the complexity of the scheduling problem. A relatively small problem to determine the timings of 154 trains for one hour of the Hong Kong MTR system involves over 40,000 constraints and 30,000 variables (including over 10,000 binary variables) [3].

2 Timetabling

Given the set of lines and overall frequencies, the timetabling problem determines the exact time of despatch from each terminus for each line. For regional trains of lower frequency, the preference of passengers is for periodic timetables (easier to remember), although often the frequencies are different for rush hour and/or weekends.

One useful mathematical model for designing periodic schedules is the *periodic event scheduling problem (PESP)* introduced by Serafini and Ukovich [1]. In this problem, we are given a period T and a set of events V , where an event represents the arrival/departure of a directed train line at a given station, and a set of constraints A . Each constraint $a = (i, j)$ imposes upper (u_a) and lower bounds (ℓ_a) on the time assigned to a pair of events. A feasible solution to PESP is an



© Olesya Kuznetsova | Shutterstock.com

3 Equipment and crew scheduling

A given timetable can be considered as a number of trips (each with a given originating location and time, given duration and ending location and time). The determination of which equipment (locomotives and coaches) and crew to operate these trips is the equipment and crew scheduling problem. A simple version of the equipment scheduling problem, where we assume that all trains are identical and originate from the same depot, can be modelled using a graph. Each trip, e.g. in a weekly schedule, is represented as a node of a directed graph. There will be an arc from node A to node B if the same train can be used to operate these trips sequentially; for example, when the destination terminal of trip A is the origin terminal of trip B and trip B starts after the end of trip A, or it is possible for the train to make a dead-head trip (carrying no passengers) from the destination terminal of trip A to the origin terminal of trip B in time to operate trip B. The graph will also have a start (S) node representing the depot location at the beginning of the week and a terminate (T) node representing the depot location at the end of the week. See the example in Figure 2.

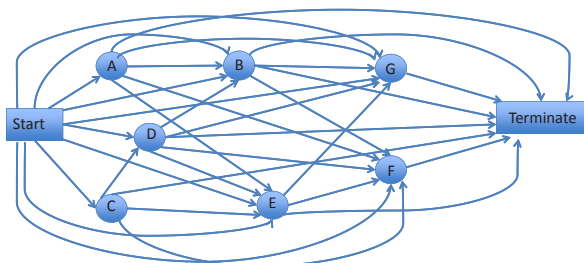


Figure 2

Since the locomotives and trains are expensive equipment, a typical objective is to minimise the number of trains needed to operate the given timetable. Thus the problem is to find the minimum number of S–T paths such that every trip node is incident to exactly one path.

If different train configurations can be used to operate the same trip, then the problem becomes more complex. Each node would then represent the operation of a given trip by a specific train configuration, and there is an arc from node A to node B if it is possible to operate trip B after trip A using the *same* train configuration. In this version of the problem, a given trip corresponds to a subset of nodes, representing the operation of the trip using different train configurations. Thus, not every node needs to be covered by a path, but exactly one node among the subset of nodes corresponding to a trip must be covered. This leads to a network flow problem on hypergraphs [4].

Table 1: Example of trips to be covered in Figure 2.

Trip	Origin	Departure	Destination	Arrival
A	Ambridge	6:00	Bedrock	7:00
B	Bedrock	7:18	Hogsmeade	8:38
C	Coronation St.	5:20	Stepford	6:05
D	Stepford	6:22	Bedrock	7:02
E	Emerald City	8:00	Gotham	9:00
F	Hogsmeade	9:15	Freedonia	10:10
G	Gotham	9:25	Ambridge	10:15

After the train configurations have been assigned, crew must be assigned to the operation of trips. The framework for assignment of crew to trips is similar to that of equipment scheduling, but there are many more constraints. There are many labour rules to be considered regarding shift lengths, meal breaks and rest breaks, required days off during a week/month, rest period after night-shift work, etc. Unlike trains/locomotives that need not return to the depot until many days later, it is desirable that a crew start and end each day in the same depot; this means that the trips obtained from the equipment scheduling problem might be split into segments to allow for a change of crew between segments.

Moreover, not all crew have the skills to operate all types of trains. This often leads to an explosion in the size and complexity of the problem. The crew scheduling problem is typically solved as a set-covering problem using a column-generation approach, where each column represents a feasible duty for the crew, and each row (constraint) represents a trip to be covered. A duty is a sequence of trips worked by the crew that respects the skills compatibility of the crew and equipment, and obeys labour rules [5].

4 Disruption management

We all wish railway systems would run like clockwork (and they mostly do), but accidents still happen. Vehicles break down, crew may be absent, signalling or power systems malfunction, and weather and traffic conditions also cause delays. Recovery of the system after a disruption is a key aspect of railway operations [6]. In case of a disruption where a part of the railway network becomes inoperable, decisions and adjustments have to be made in real time, often with limited information as to the extent and duration of the disruption. If a part of the tracks are blocked or inoperable, some lines would need to be re-routed or cut short, trains may be cancelled and/or the timetable adjusted. Vehicles and crew will need to be redeployed to operate the trips for the adjusted timetable. Passengers react to news of the disruption by choosing alternative routes or modes of transport, thus estimating the demand for the disrupted network and ensuring sufficient capacity for the modified demand is particularly difficult [7]. A key aspect of the disruption management of railway systems is the real-time reallocation of tracks to accommodate the modified timetable and routings of the trains. The carrying capacity of the system is reduced, yet safety distances and time gaps must be maintained. Both passengers and crew have to make modifications to their habitual routine, often leading to confusion and further delay. Particularly at bottleneck areas such as major stations, ensuring that the modified system can be run with minimal changes to the standard plan is a very difficult task [8].

4.1 Urban tram systems

Unlike most rail systems, some urban rail transit systems do not run on dedicated tracks but must share the space with vehicular traffic on the road. One such system is the Hong Kong Tramways system that operates double-decker trams serving a densely populated area of Hong Kong. In the congested central business district, the trams are often blocked by other vehicles; disruption management becomes a continual task! The system serves the major business and commercial districts as well as key residential areas on Hong Kong Island, thus the demand pattern varies quite a lot over the course of the day. Because of traffic congestion, the travel time of the trams are also highly stochastic. The



Hong Kong trams share space with other road users, including pedestrians, buses and private cars.
© Gelia | Dreamstime.com

service is a high-frequency service, so passengers are not aware of the exact timings of the timetable, although they *expect* a headway of 1 or 2 minutes. Because, the stations are on the road, long waits would lead to balking, so maintaining a high frequency service is very important. When a tram reaches a terminal station, it is possible to reassign the tram and driver to alternative routes – different from their scheduled duties. We have been developing a decision-support system to support the real-time management of the tram system. Our system is based on a look-ahead model that tries to find the best set of reassignments for all trams and drivers that will be arriving at some terminal within the next time period. This look-ahead model is solved repeatedly on a rolling horizon basis (using the latest updated information) thus allowing the controllers to dynamically revise the schedule in real time. For all trams arriving at a terminal within the planning period, we consider all possible subsequent schedules that could be assigned to the tram and evaluate the cost (in terms of overtime, demand coverage, meal-break delays, etc.). Using a matching-based model, reassignment of routes for all trams arriving at the terminals within the planning period is optimised. The objective is to maximise the route frequencies in order to provide good service to passengers, and minimise the violation of staff regulations (meal-break delays and overtime) taking stochastic time-dependent travel-time uncertainties into account [9].

5 Future trends

As mentioned, because railway design and operations are so complex, the planning has traditionally been done in stages. Experience and research studies have shown that this decomposition of the problem may lead to suboptimal plans. With advances in mathematical modelling, algorithm development and computational capabilities, researchers and practitioners have been investigating integrated planning models, which would enable better overall service quality and efficiency. Models for the combined planning of equipment and crew are being developed. Another direction of research is on models that take uncertainties into account, so that the operational plans are more robust and can better

absorb minor disturbances. A pressing research challenge is in real-time disruption management. The availability of real-time data (e.g. from social media on immediate passenger demand, from sensors on accurate location of trains) may lead to new models and systems for railway planning.

REFERENCES

- 1 Serafini, P. and Ukovich, W. (1989) A mathematical model for periodic scheduling problems, *SIAM J. Discret. Math.*, vol. 2, no. 4, pp. 550–581.
- 2 Liebchen, C. (2008) The first optimized railway timetable in practice, *Transp. Sci.*, vol. 42, no. 4, pp. 420–435.
- 3 Wong, R.C.W., Yuen, T.W.Y., Fung, K.W. and Leung, J.M.Y. (2008) Optimising timetable synchronisation for rail mass transit, *Transp. Sci.*, vol. 42, pp. 57–69.
- 4 Borndörfer, R., Reuther, M., Schlechte, T. and Weider, S. (2011) A hypergraph model for railway vehicle rotation planning, *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization and Systems (ATMOS)*, eds Caprara, A. and Kontogiannis, S., pp. 146–155.
- 5 Kroon, L., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., Schrijver, A., Steenbeek, A. and Ybema, R. (2009) The new Dutch timetable: the OR revolution, *Interfaces*, vol. 39, no. 1, pp. 6–17.
- 6 Cacchiani, V., Husiman, D., Kidd, M., Kroon, L., Toth, P., Veelen-turf, L. and Wagenaar, J. (2014) An overview of recovery models and algorithms for real-time railway rescheduling, *Transp. Res. Part B*, vol. 63, pp. 15–37.
- 7 Schmidt, M. and Schöbel, A. (2015) The complexity of integrating passenger routing decisions in public transportation models, *Networks*, vol. 65, no. 3, pp. 228–243.
- 8 Caimi, G., Fuchsberger, M., Laumanns, M. and Lüthi, M. (2012) A model predictive control approach for discrete-time rescheduling in complex central railway station areas, *Comput. Oper. Res.*, vol. 39, pp. 2578–2593.
- 9 Leung, J.M.Y., Lai, D.S.W., Kuo, Y.-H. and Cheung, H.K.F. (2016) Real-time integrated re-scheduling for public transit, *3rd International Conference on Railway Technology: Research, Development and Maintenance*, ed Pombo, J., Civil-Comp Press, Scotland.