Rethinking train scheduling to improve network capacity management

Carl Van Dyke  TransNetOpt
Managing Director
Senior Advisor – Oliver Wyman
N.A. freight railroads often schedule/dispatch trains in a manner that:
  - Ignores network line capacity
  - Does not fully take into account the competing needs of each line of business (manifest, intermodal, unit train, etc.)
  - Has only a loose understanding of scheduling requirements

The introduction of optimized movement planners does not resolve this fundamental, network level, capacity allocation problem, but in fact may make matters worse by attempting to move more trains over a line than can be reliably handled.

Strategies in other parts of the world take a more structured approach that appears to deliver better capacity utilization, tighter schedule adherence, and greater reliability.
  - These strategies may not be feasible in the N.A. railroad business environment

The N.A. industry needs to rethink how it schedules its trains and manages its network capacity.
  - One possibility is to use a mid-tier capacity allocation process that acts as a layer between the initial train scheduling process and the dispatching process.
  - Such a process would be focused on managing the trains to ensure the fluidity of the network and minimize the occurrence of congestion problems and the attendant network and service impacts.
  - Result would be to create a continuously maintained set of real time “supervisory” train schedules.
Presentation Overview

- Global Perspective on Network Utilization Strategies
- What is Network Capacity?
- Reinventing How Schedules and Capacity are Managed
North American railroads are not line capacity leaders and operate at a capacity discount relative to foreign railways

In Europe, all trains have a plan: Scheduled trains have their own slot, while unit trains are allowed to run in pre-planned, but otherwise unassigned slots.

For N.A., could a strategy of setting the timing of trains 24 to 72 hours in advance against line capacities allow railways to better utilize their networks and improve train/asset velocity?

Source: UIC International Railway Statistics 2012; Association of American Railroads; and Oliver Wyman analysis
European slotting strategy

Reserves day-of-year specific, time-of-day specific paths (or “slots”) for trains over every line segment

- A train operating on path A-B-C-D may have 3 slots: A-B, B-C, C-D
- 30% of freights in Germany, and 100% of Freightliner trains in the UK are scheduled <36 hours before departure, but the patterns are predictable enough to usually fit them into predefined slots

Operator must commit to tight departure time and arrival time windows, and powering trains to maintain their planned track speed, which is largely uniform for all trains on a line

- Pay a premium if train is faster or slower than the “standard” speed
- Failure to have train ready within a tight time window (+/- 1 minute) can mean loss of path

Planning process is 18 to 21 months out. Plan is largely frozen for 12 months

In Germany, slots planned 1+ years in advance pay $2.20 to $6.00 per train km

- Price adjusted for speed (slower or faster trains pay more), locals might pay less
- Short term changes to plan/dynamic slots can cost double
- Ability to reserve extra slots, with substantial cancelation fees

Each operator must cooperate with central network authority to get slots

- Network authority must fairly assign slots to multiple operators

Changes to trains operated and “ad hoc” unit trains either use slots already owned by operator, or must be “purchased” on a dynamic basis (generally from track authority, not other train operating companies)

Slotting reduces irregular trains which are seen as capacity destroyers
Higher network performance requires technologies

Technology and discipline must be used to sustain capacity and increase the robustness of the network.

- More technology means that Germany can move much higher volumes through a bottleneck...

- In hump yards, short dwell times of 4 to 8 hours and retarder technology allow yards to sort everything from autos to containers to chemicals and meet high service levels.

Typical German Yard
Higher network performance requires technologies

Technology and discipline must be used to sustain capacity and increase the robustness of the network

...A typical North American yard has little to no comparable technology

Typical North American Yard
Designing for throughput management

Building strategically located train staging capacity allows trains to be released into each network segment on a controlled basis.

3 Trains waiting for paths (slots) – while also taking care of scheduled crew changes. Note signal masts controlling staging process. Drivers arrived to locomotives ~15 minutes prior to departure.

This type of approach is not likely to be adopted in the N.A. environment – so what are the alternatives?
What is Network Capacity?
What is rail network capacity?

High level view is that network capacity is the total amount of freight and passengers that can be moved through the network

- Networks can have multiple routes, which means that the specific path taken by each shipment and passenger is an integral part of the capacity puzzle

Many different capacities must be considered in determining the overall capacity of a network

- The key elements of capacity can be summarized as lines, yards, resources, and operating strategy

All of these elements play a critical role in determining the total capacity:

- The number of trains that can traverse a rail line is a core element of capacity – but is a function of many factors such as the line configuration, signaling, type of trains operated, their relative speeds and sizes, and the operating strategy
- Yards can influence capacity, particularly for “carload” traffic, and impact line capacity based on the space available to “land” trains traversing the line segments
- Resource availability is also critical (number of railcars available to move shipments, locomotive fleet size available to move trains, access to qualified engineers/drivers, etc.)

As one increases the number of elements considered, the complexity/practicality of the assessment process becomes significantly more challenging
The maximum capacity of a rail line depends on many factors including the track configuration, the speed and size of the trains being operated, and the signaling system.

- For example, for a typical single track line with passing sidings, the “nominal” capacity per day is the maximum running time between sidings in minutes divided into the number of minutes in a day (1440).

\[
\text{Capacity} = \frac{1440}{\max(30,45)} = \frac{1440}{45} = 32 \text{ trains/day}
\]

(assumes equal numbers of opposing train movements perfectly interleaved)

However, if some trains are faster or slower than others, this will directly impact effective capacity, as would operating strategies such as “fleeting”.

- For a double track line, maximum capacity is roughly equal to the minimum headway between trains divided into 1440 and then multiplied by 2.
Many other factors impact the capacity consumption of an individual train relative to others operating over a line, such as speed/performance, delay tolerances, ability to fit in sidings, need for station stops, etc.

Furthermore, the maximum capacity can only be achieved if everything goes perfectly
  - Practical capacity is typically 60% to 80% of the maximum

This practical capacity can be further adjusted by factors related to the ability of the yards to “launch” and “land” trains (which may depend on train type)

Overall, one can thus characterize the capacity of each part of a network in terms of the “practical” number of trains that can be operated in each direction per a period of time (such as per hour, or per four hour block)
  - Capacity also needs to be defined in terms of several threshold points as depicted in the bar graph

Normalized rail line train throughput capacity
Question becomes how many trains can traverse an extended rail line/network segment and how will trains with differing objectives interact?

A corridor such as New York to Chicago, or the “Sunset Corridor,” must also account for other movements sharing the same track segments.

From a network perspective, alternative routings should be taken into account.

Routing choices and differences in train density, line configuration, and facilities all impact the effective capacity of the network and how trains should be scheduled.
Network train throughput capacity

• From a network perspective the need for trains to enter yards must be taken into account

• Trains enter and leave yards for many reasons:
  – Originations
  – Terminations
  – Crew changes
  – Pick-up/set-off cars
  – Fueling
  – Inspections
  – Holding for on-ward line capacity

• Thus network throughput capacity is related to a mix of traditional line segment capacity and required yard capacity
Characterizing the throughput of a network

• Most railroads have a sense of the practical or working capacity of each line segment to handle train
  – Such estimates can always be refined using computer models, field observations, and other analytical approaches.

• The same can be done for yards in terms of the number of trains they can handle per time period for originating, terminating, and intermediate yard calls.

• In general these capacities reflect a “normal” train for that network segment
  – Some trains may consume more line or yard capacity due to their length or performance attributes, and some could consume less

• It is likely that one could develop a rough “rule of thumb” value for the relative capacity consumption for each train type, for example:
  – A 100 car manifest train might be your “normal” train
  – A large, heavy, slow coal train might consume 1.5 times the capacity of the normal train, a local might be 2.5, while a short, fast intermodal might only be 0.75

• Using a basic approach like this allows one to look at the trains traversing a network segment and understand in rough terms their consumption of the segment’s capacity
  – Using this framework, one can also see where train timings (or routings) should be adjusted to better balance the use of the network routes and avoid congestion issues

• The Intelligent Train Scheduler model used by a number of railroads reflects in part this type of approach
If segment D-J is closed, we will have a need to move 15+23+18 or 56 trains/day over a combination of segment C-G-H and F-L
  – Assume these two segments can at most handle 50 trains/day
Which 6 trains worth of business do we not move, or try to handle in a different manner?
  – Only impact customers that used the D-J segment?
  – Share the pain over all customers?
  – Prioritize the traffic based on near term profitability?
  – Consider the strategic importance of each customer in making the decision?
Or can we increase capacities in the short term?

Capacity allocation decisions must be made anytime volumes exceed capacity, even when the network is fully functional.
Reinventing How Schedules and Capacity are Managed
All Class I railroads maintain a set of “dated” or “operational” train schedules going out 7 to 14 days into the future

- This “real time” train database is used to manage and track all train movements, plan yard work, set performance goals, call crews, and set customer expectations
- No train can be moved that is not in this database
- Unfortunately, the train timing in these databases is strictly “notional” and has limited correlation to reality

Each schedule is largely created independently, without any consideration of other trains that will be operated based on an arbitrary “launch date and time.”

- That is, the person setting the schedule picks a starting date/time for the train, and the rest of the schedule is set using nominal running times without consideration of any other factors or network effects
Supervisory schedule concept

Using a network capacity model/tool, one could conceive of a process where a capacity check was done when a train is proposed, and the timing of the train was adjusted so that there would be capacity for the train on each network segment the train was to traverse

- These timing adjustments would include holding of the train at origin, as well as potentially holding the train at intermediate points (staging) when the train might otherwise enter a network segment when it was at or above capacity

- Decisions would reflect line capacity, staging capacity, relative priorities, and target arrival times (want dates/times)

Idea behind a set of “supervisory” train schedules is to set the “dated” schedules based on the availability of network capacity as described above

- In effect, the supervisory schedules would meter the movement of trains through the network, ensuring network fluidity, schedule achievability, and dispatcher sanity
Setting the supervisory schedules

**Base Schedules**
- Train routing
- Nominal timing (schedule offsets) for each train when run
- Planned work events and stops for crew changes, etc.
- For “fixed schedule” (recurring) trains, specific scheduled times partially validated against network capacity including other trains
- Tolerances on any targeted arrival times

**Recurring Trains**
- Auto-added to supervisory schedules
- Modest adjustments to schedules to achieve capacity feasibility
- Addition might cause adjustments to other trains
- Generally placed in database before ad hoc trains, with higher priority than ad hoc trains
- Alert if cannot be accommodated

**Ad Hoc Trains**
- Manually added trains, using base schedule as template
- User specifies scheduling goals – priority, departure window, arrival goal (want date/time)
- Schedule adjusted (significantly?) to fit available capacity
- May be more subject to adjustment later as additional trains added

**Supervisory Schedules**
- Continuously updated database of all date specific trains being operated or anticipated to be operated
- Reflects current status of all trains including location and timing
- Schedules set in manner to ensure feasibility against available capacity
- Staging events (holds for capacity) reflected in schedules where required

**Schedule Event Manager**
- Receives and evaluates each train movement event and origin train release
- Based on events and releases, adjusts schedules to continuously maintain flow against capacity
- Generates alerts for trains that will be delayed to maintain network fluidity

---

© 2017 TransNetOpt & Oliver Wyman
Solution strategies

Creating and maintaining the supervisory schedule layer can be done today with relatively simple solution strategies

– Characterize each line segment in terms of train throughput along lines discussed earlier

– Use a network level, basic flow model to set approximate schedules respecting limits on maximum trains per segment per hour (likely implemented with a simple, first come, first served heuristic approach over a time-space network)

– Refine schedules using a tool that takes into account the attributes of the trains and lines such as the Intelligent Train Scheduler model

Above tools would need to be extended to take a number of factors into account:

– Need for staging to meter flow across line segments, and tracking basic capacity for staging trains

– Track target arrival dates/times (or want dates/times) and factor this into how aggressively trains will be moved

– Could move towards optimization and more accurate capacity assessment over time as experience is gained

Leave the detailed dispatching and management of trains on each line segment to the dispatchers and associated dispatching systems
Impediments & benefits

Impediments

– Siloed scheduling organizations (manifest, intermodal, automotive, coal, grain, other unit)
– Need to change historic practices, and introduce more discipline in the scheduling of trains including reliance on analytics to make decisions/
– Railroad culture / management
– Technology development risks / costs
– Excess focus on movement planners / dispatch optimizers as solving this problem

Benefits

– Reduction in delays / improvement in on-time performance
– Improved train velocity
– Reduced crew costs / tighter call times for crew changes
– Better customer service / ETA estimates
– Improved ability of yards and customers to plan workloads, increasing asset velocity