2012 RAS Problem Solving Competition

Movement Planner Algorithm Design For Dispatching On Multi-Track Territories

DISCLAIMER: The problem presented here exemplifies one of several opportunities for Operations Research application in the Railway industry. We have simplified a real-life problem for this competition. More general problems and related literature is available on the competition web-site under the section Related Literature. Best of luck for the competition.

informs ras
Railway Applications Section

This year's RAS Problem Competition addresses a real life problem that the railroad industry has been trying to find a benchmark solution for decades. The problem is scheduling the movement of trains optimally over a detailed track network in such a way that the trains do not collide, honor operational constraints and achieve maximum overall efficiency (or minimum overall delay). The problem is generally known as the "Dispatching" or "Meet-Pass Planning" problem, the tools used to address this problem in real-time are generally called "Movement Planners," and the tools used to explore this problem on a planning basis are called "Line Capacity Models" or "Dispatching Models." While many solutions exist for the planning mode of this problem, only after decades of work are a few solutions for the real-time variation now starting to enter operational use.

Numerous publications address this problem class. These publications can be searched using keywords such as Meet-Pass Planning or Train Timetabling. Some of the most prominent papers in this area are listed in the references section of this report. Please note, that the version of the problem introduced in this year's competition has some real life challenges that have not been fully tackled in the literature.

This document briefly explains the dispatching problem, and then the process by which trains are issued "movement authorities" also called "track authorities". Next, relevant railroad jargon is introduced along with the operational rules (constraints) and the objective function components. The feasibility of the solution and the objective function components will be used to judge the competition entries to ensure that a systematic approach is used to evaluate each entry. It is not enough to "solve" the problem; the best solution needs to solve the problem very quickly, so the solution time will also be a factor in the judging. In addressing the real-time movement planner requirements, this problem must be solved continuously in a matter of seconds or minutes to be effective. The contest judges will thus be looking at this aspect of the solutions in addition to the objective function value and any constraint violations. The detailed judging criterions are mentioned on the competition webpage.

Please read the problem statement and "toy problem" description carefully. One must look at this document in its entirety to understand all of the constraints that must be observed.

Introduction

To maintain safe flow of train traffic on the railroads, railroad companies employ people with very dynamic decision making skills, called dispatchers. Dispatchers have full authority to decide which trains are assigned the right to use which track segments for particular time periods. The authority given to the dispatcher typically applies from a specific track location to another (generally called mileposts), and everything in between these mileposts constitutes the "Dispatching Territory" for a specific dispatcher. These territories can consist of single-track segments along with passing sidings, or they can be formed by a combination of single-track segments and multi-track segments with sidings and cross-overs (locations where trains can move from one main track to another). As the names imply, single-track segments allow only one train to have movement in one direction in a given period of time, whereas a multi-track segment lets more than one train to move between the beginning and ending mileposts by utilizing separate tracks. When two trains are to meet each other on a single-track segment (see Figure

1), one of the trains must take a siding prior to arriving to this segment. Arcs 2-4 and 6-7 are called the switch tracks, which allow trains to move to the siding from the main track and move to the main track from the siding, respectively. Similarly, on the multi-track stretch (see Figure 2), if the number of trains entering this stretch is more than the number of parallel tracks, one or more of the trains must stop at sidings to avoid a collision.

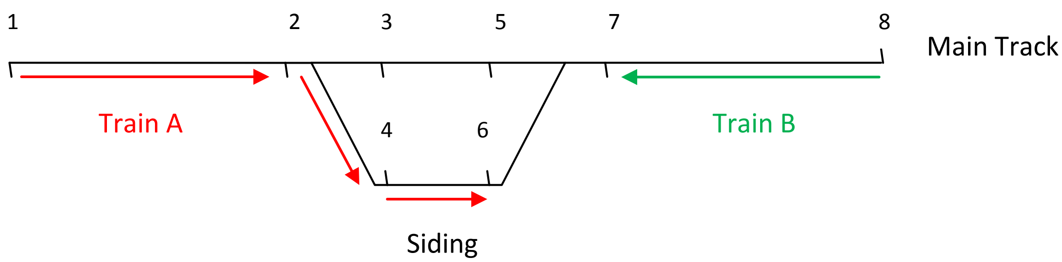

Figure 1: Two trains meeting on single-track.

In multi-track territories, trains can move between mainline tracks using crossover tracks (see the arcs between nodes 2-17, 5-18, 10-27, and 13-28 in Figure 2). For instance, in a situation where two trains are both on Main track 1 and close to each other, crossover tracks 10-27 and 13-28 enable these two trains to run concurrently between nodes 10 and 13 on Main Track 1 and between 26-29 on Main Track 2. Nodes 10 and 26 have the same milepost, as do the nodes 13 and 29. That is why this portion of the territory is called Multi-track, because there is more than one way to travel between mileposts. These parallel tracks can be used in the same direction or in opposite directions.

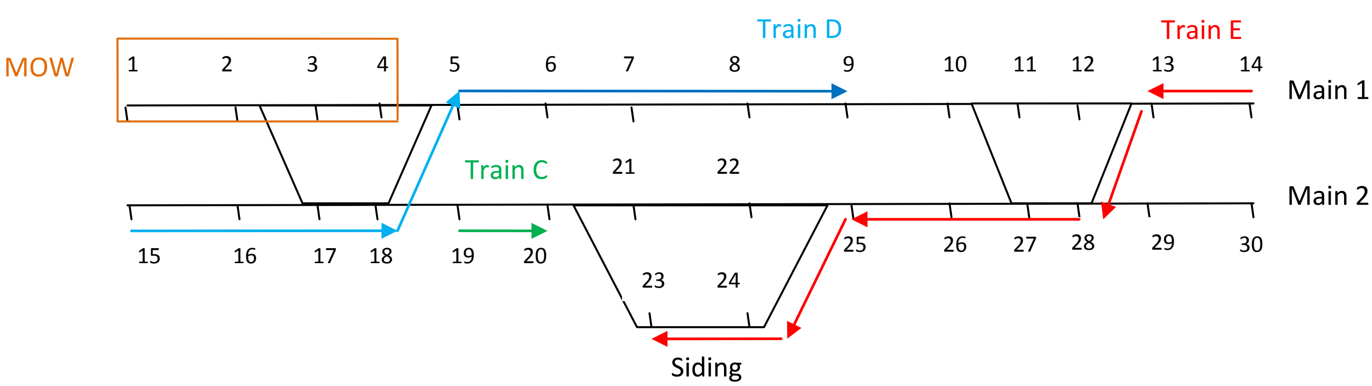

Figure 2: Three trains meeting on double-track.

In Figure 2, Train D and Train E are moving toward each other, and the dispatcher instructs Train E to use the crossover track 13-28, so that Train E does not collide with Train D and to continue on Main Track 2. Furthermore, since there is another train (Train C) coming toward Train E now that Train E is on Main Track 2, the dispatcher assigns Train E to take the siding until Train C clears the arc 20-21. For Train E to continue on Main Track 2 (start to occupy track 23-20), the rear end of Train C must clear from the track on arc 20-21. The decision as to which train takes the siding is the result of a multiple decision criteria, which will be expressed using a multi-criteria objective function described in detail later in this problem statement.

Figure 2, also illustrates the notion of MOW (Maintenance of Way) windows. Due to repairs or inspection activities on the rail tracks between nodes 1 and 4, this portion of the railroad might be unavailable to train traffic for a specific amount of time. Therefore, Train D moves on Main Track 2 and then takes crossover 18-5 to bypass the MOW track outage.

Trains on railroads have different speed profiles. The speed is dependent on six main factors among several others: (1) The throttle setting, (2) the total horsepower of the locomotives, (3) the grade of the territory (being downhill or uphill), (4) the tonnage of the train, (5) aerodynamic drag resistance and (6) rolling resistance. The speeds of the trains are also impacted by the speed limits assigned to each segment of track. As a result, as trains move over a line, their speeds vary based on the above factors, and the time it takes for the trains to traverse a section of the railroad depends on acceleration and deceleration due to changes in speed limits, and the need to start and stop for other trains. To simplify the complex calculations required to determine precise train speeds under various situations, for the purposes of this problem, we will treat each train as having a fixed (maximum) speed based on its train class (A, B, C, D, E etc.), the track type (Main 1, Main 2, Switch track, Crossovers, Sidings etc.) and the direction of movement (Eastbound or Westbound).

On a daily basis, dispatchers face the challenge of dispatching trains on both single-track and multi-track territories. Single-track dispatching is more difficult for a dispatcher to handle; it requires more careful planning since there is only one track for the trains to move on in either direction. From an algorithm design standpoint, the multi-track dispatching is a more challenging problem because of complexity, number of paths available for train movements, restrictions, etc. While a track closure (outage) for MOW or other reasons on a single-track territory completely shuts down the network for a predetermined amount of time, the multi-track territory can still allow trains flow towards their destinations using the parallel track to the one that is closed (assuming the parallel track is still open).

The Movement Planner Problem can be described by the following: Given a dispatching territory, a dispatcher's task is to give track authorities to trains to maximize the overall system efficiency of the trains while abiding by a number of business rules. This is done by deciding which train takes the siding when two trains traveling in opposite directions meet each other, or when a faster train is to overtake the leading slower train. The dispatcher must also decide on the departure and hold times at the terminals where trains originate, terminate or stop for work events (picking up cars, setting out cars, inspection, fueling, changing crews, etc.).

In this problem setting, the territory is comprised of sections of single and double tracks. Both the single-track and the double-track sections have sidings where the non-favored trains can be held until the next segment en route is available. Terminals exist at both the west and the east ends of the territory, and there are no other terminals inside the territory, all nodes along the route of the trains are pass-through, where no work events occur. Crew hours of service are not a constraint in this problem.

Initial Assumptions and Preliminary Information

- Instantaneous acceleration to track speed from a stop and deceleration to halt from full speed are assumed. A train **must** move at the prescribed speed for each track segment or completely stopped. Train movement at a speed other than prescribed speed is not allowed.

- The territory utilizes a simplified signaling or control system. Normally, the signal system is used to space trains apart based on various safety rules and train separation rules. The signals appear at control points, which are places where trains can proceed on one of the tracks ahead (based on the alignment of the rail switch). In addition to these control points, other intermediate points on the network may also have signals to control use of subsections of a track by the trains. These signals notify the engineer whether the proceeding track is available to continue or not. However, there are neither intermediate signals nor control point signals in this problem setting. Simplified occupancy rules are described in the next bullet point.

- Only one train can occupy the track between two consecutive nodes at a time (move or wait) for safety reasons. That means, in a case where two trains are moving in the same or opposite directions, the second train cannot enter a track arc occupied by the first train until the rear end of the first train clears and exits this arc.

- Trains can only move in their pre-specified direction. Backing-up is not allowed. For example, if an eastbound train entered a siding, it must leave the siding from the east end, and not from where it entered the siding.

- Deadlocks, wherein the continued movement of the trains can only be achieved by moving some of the trains in the reverse direction, must be avoided. Figure 3 is an example of a deadlock situation. Train A needs to reverse in order for Train B to take the siding, so that Train C and then Train A can continue eastbound.

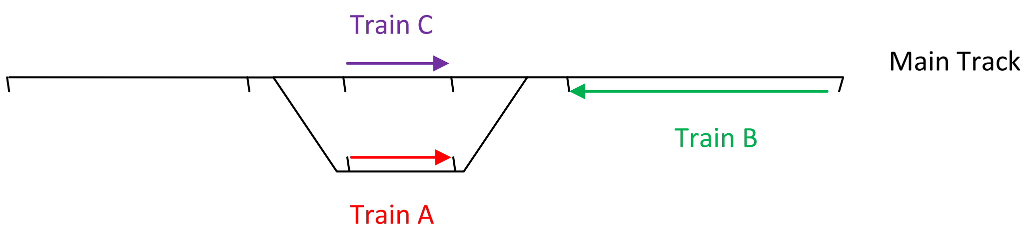

Figure 3: Deadlock

Definitions

Train entry and exit time: Train is assumed to enter a track arc when the leading end of train enters into the arc. Train is assumed to leave a track arc when entire train length exited the arc.

Train Delay: Train delay is defined as the time lost while the train is stopped at a terminal or anywhere on the territory because of a MOW or for meet/pass for other trains.

Schedule Adherence: Regarding schedule adherence, there are two classes of freight trains. There will be no passenger trains in the data sets.

- Schedule Adherence (SA) trains: These trains' movements are expected to conform to a predetermined schedule. Any deviation from the schedule should be avoided unless it is not impacting the average system velocity significantly. Train types A, B, C and D are examples of such trains.

- Non-Schedule Adherence (NSA) trains: Although their expected entry times are known, these trains' movements do not necessarily need to abide by their schedules. These trains have lower priority when it comes to track allocation decisions. Train types E, and F are examples of such trains.

Schedule adherence is based on train entry time at a node.

Train priorities: A good guideline for assigning trains to tracks can be given by the following relationship among trains:

A > B > C > D > E > F

This priority relationship is derived mostly based on their velocities, but also based on their schedule adherence attributes.

Station: A station can be construed as the collection of nodes (e.g., nodes 2 through 7 can together be called a station) in Figure 1

Meet-Pass event: A Meet-Pass event can take place under 2 conditions:

a- Meet: Two trains moving in opposite directions meet each other on the same main track at a station where there is a siding or there is a crossover to the other main track. During a meet event, if the first train arriving to this station takes the siding, it will have to wait until the other train arrives and moves past the first train, clearing the main track. If the first train arrives and waits on the main track due to some operational rules (given below), then the second train must take the siding and move on.

b- Pass: Two trains moving in the same direction, where the following train is faster than the leading train. In this case the dispatcher may elect to have the following train overtake the leading train in what is called a "pass" event. Similar to a meet, in single track territory, the first train would pull into a siding and wait until the following train passes it on the main track. In the double-track portion of the territory, trains are allowed to use crossovers to move to the other parallel track and execute the overtake or pass, thus avoiding a possible collision.

MOW Windows: Due to government regulations, for the safety of the employees and communities, the rail tracks and signals on the network must be regularly inspected and repaired when necessary. A closure of a track for inspection or repair is called a Maintenance of Way (MOW) track window, where the "window" refers to a specific period of time during which the track cannot be used. Although, MOW track windows are intended to be placed at times when there is very little traffic, sometimes, a broken rail or another issue requires immediate attention and closure of a track segment.

When a MOW window occurs, the dispatcher needs to close that portion of the railroad to train traffic to keep the rail gangs (repair employees) safe. A MOW window is given with two attributes: space (from which node to which node) and time (beginning minute to ending minute). When a MOW window is placed on a single-track on the main line where there is no alternative track (like a siding) to go around; all the trains that are destined to traverse the closed track have to be held somewhere prior to arriving at the impacted track during its outage window. The holding places could be the terminals at both ends of the territory, the sidings and the main tracks previous to the closed track. On a multi-track stretch, trains can use crossovers to change tracks so that the trains to continue to operate past the section of closed track by moving on the parallel track.

The dispatcher places the trains on the main track segments, sidings or hold them at the terminals in such a way that, when the MOW window is about to be over, the trains having the highest priority, and critical schedule adherence obligations are least impacted by the delay.

Terminal Want Times: The territory includes two terminals, one on the west end and one on the east end. Terminals have finite yard track capacities to handle work-events. When there is heavy inbound or outbound traffic at a terminal, the terminals may ask the dispatcher to delay entry of some trains into the terminal in order to accept them in their yards when they believe they have the capacity to do so. A Terminal Want Time represents the preferred time for a train to arrive at a terminal. While it is desirable to abide by the Want Times, they are **not** hard constraints. It is more important to run a railroad without deadlocks. When a terminal issues a Terminal Want time for a specific train, it is generally appreciated if this particular train arrives within the next three hours of this Terminal Want time. Moreover, it is also undesired to make this train arrive earlier than one hour before the Terminal Want time. This preference gives a four hour <u>preferred</u> time period for the train to arrive at that terminal.

Terminal want time is based on train entry time at a node.

Operational Rules (Constraints)

- Heavy trains: Trains over 100 TOB (tons per operative brake) are considered to be heavy. Heavy trains can belong to any one of the train types, A through F. In cases where a heavy train is undergoing a meet-pass event and the second train is of a Non-Schedule Adherence type, the heavy train **must** hold the main track, and the NSA train takes the siding. If the second train in consideration is of a Schedule Adherence type, regardless of the type of the heavy train, the decision on which train uses the siding is based on the overall system efficiency.

- Long trains: Long trains cannot use the sidings which are shorter than the train's length. The siding arcs are marked with an 'S' in the data files with their corresponding lengths. The length of the switch tracks cannot be added to the siding length to determine whether the train fits this particular siding or not. For instance, in Figure 1, the train length must be compared to the length of arc 4-6, not the sum of the lengths of 2-4, 4-6 and 6-7. This long train must wait at the corresponding main track (arc 3-5) regardless of which train type is incoming.

- Inhalation Hazard Trains: Inhalation Hazard (IH) trains are forbidden to occupy the sidings to ensure public safety.

- Schedule Adherence Trains: When a dispatcher is to make a decision among the same type trains during meet-pass planning, he checks whether each is on time or behind schedule. Same type Schedule Adherence trains are assumed to have the same velocity. A Schedule Adherence train is considered to be "hot" when it is on-time, ahead of its schedule or it is less than 2 hours late. If the train is more than 2 hours late, then it is considered a lost cause and schedule adherence becomes less of a priority than ensuring that other trains stay on schedule. The following table only suggests what to do during such meet-pass events, assuming there are no other constrains imposed such as terminal want times, unpreferred track occupancy, etc.:

  Table 1: Priority Relationship Suggestions between Same Type Schedule Adherence Trains.

	A1	A2	Decision
Case 1	15	25	System efficiency
Case 2	-60	-90	A1 waits
Case 3	-60	-120	A2 waits
Case 4	-60	-180	A2 waits
Case 5	-140	-180	System efficiency
Case 6	30	-90	A1 waits

  A decision labeled "System efficiency" means that the dispatcher takes an action that will maximize overall system efficiency. Positive numbers in Table 1 indicate that the trains are ahead of their schedule, i.e., they arrived at that location that many minutes before their scheduled arrival times, and negative numbers denote that they are behind schedule.

The aim of this constraint is to keep the trains "hot". When both of the trains can maintain their status as "hot" (Case 1), then the decision is based on overall system efficiency. This is also the case when both of them are already behind their schedules more than 2 hours (Case 5). In Case 2 and Case 6, there is a better chance that Train A2 can stay hot if Train A1 waits. Opposite is true in cases 3 and 4, because Train A2 already lost its hot status. Table 1 only provides a rough guideline. It is given to emphasize the importance of the 2-hour limit as an incentive. The optimal solution **may** propose a different meet-pass plan due to its benefits relative to observing the 2-hour incentive.

- Unpreferred track usage: Taking the siding versus moving on the parallel track is driven by overall system efficiency and is left to the discretion of the dispatcher. In the case of multiple parallel tracks, each track will have a specified preferred direction. Trains should move on their preferred tracks instead of the available parallel track whenever possible. It is preferred that Eastbound trains use Main Track 2 and westbound trains use Main Track 1 whenever possible. The sidings and switch tracks directly connected to unpreferred main track are also unpreferred for train to occupy. In Figure 2, train E is using unpreferred track (main 28-25, switch 25-24, siding 24-23 and so on until it uses a crossover to come back to Main Track 1) after using crossover 13-28.

- Train delay: The solution must not show any delay (waiting) on a siding when there are no meet-pass events occurring. In order to relieve an impending congestion ahead, the solution cannot assign trains to sidings. A better solution would be slowing down the train prior to the congestion. This can be effectively done by letting the train traverse that arc in its normal pace, and then instantaneously stop and get delayed by the amount of time the solution suggests.

- The planning horizon is 12 hours. The penalties for Schedule Adherence and Terminal Want Times will only be activated for the events that take place within the planning horizon. Similarly, the penalties for delay time and unpreferred track usage will also be counted only for the time spent in the planning horizon.

Objective Function

The objective is to design an algorithm that provides a solution which is first and foremost feasible and deadlock-free (namely, no accidents and no trains backing up), while minimizing train delay, maximizing schedule adherence, minimizing deviance from terminal want times, and minimizing usage of tracks against their preferred direction.

Numerical relationships can be given for the objective function components, as follows:

- A 60-minute delay (regardless of being at the terminal, or anywhere on the territory) costs the railroad the amounts shown in Table 2 (due to more crew starts for the same distance travelled, dwell at the terminals etc.):

  Table 2: Relative cost impact of delay on train types

	A	B	C	D	E	F
Hourly Delay Cost	$600	$500	$400	$300	$150	$100

- A 60-minute deviance from hot Schedules costs the railroad approximately $200 (due to loss of customer goodwill). This only applies to SA type trains which are late more than 2 hours. Namely, every extra hour of tardiness beyond 2 hours costs $200.
- A 60-minute deviance from a Terminal Want Time window costs the railroad approximately $75 (due to creating undesired congestion at the terminals, hence requiring more yard crews). The cost is only incurred for the amount of deviance outside the 4-hour window.
- A 60-minute utilization of an Unpreferred Track has a penalty of $50.

Disclaimer: These values are fictional and are not the actual cost figures used by a Class I railroad company.

The total cost for system can be calculated as follows:

Total Cost = (Total Delay * Delay Penalty/Hour) +

(Schedule deviance over 2 hours for SA Trains * Penalty over 2-hour deviance/Hour) +

(Terminal Want Time deviance beyond the 4-hour window * Penalty for TWT/Hour) +

(Unpreferred Track Time * Penalty for Unpreferred Track Utilization/Hour)

The solutions from the competition participants will be evaluated based on this function considering all trains in the system during the given planning horizon.

The toy example below demonstrates how a simplified network operates.

A Toy Example:

Assume we have a single-track dispatching territory (Figure 4) with 13 nodes, where node 0 and 12 are the end terminals and there are two sidings in between. The planning horizon for this toy example is **150** minutes. Also suppose that there are 3 trains to appear in the system in the next 150 minutes. The speed, track arc length, type of track, MOW window and details about trains are defined in TOY DATA SET text file. There is a READ ME for INPUT text file which describes how to read this file.

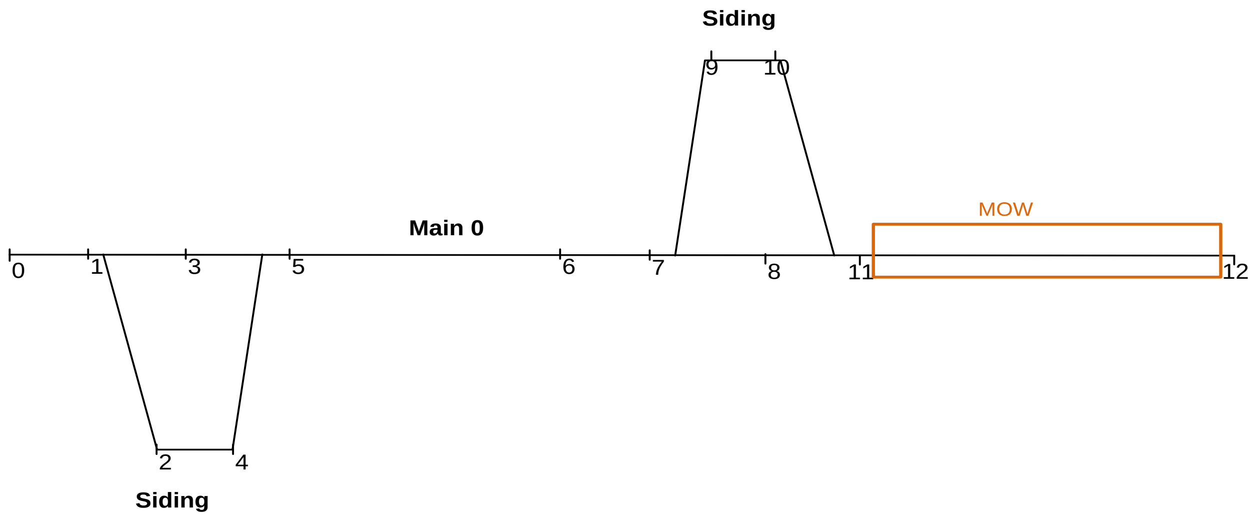

Figure 4: Toy example territory.

If there was no MOW window, and the tracks between the stations had infinite capacity, Figure 5 would show the time-space diagram that predicts the expected locations of each train at any given time. Since there is a MOW window, Train B1 cannot start moving at time 0. Moreover, since this is a single-track dispatching territory, the graph shows that Train B1 will collide with Train A1 and Train C1. It also shows that train A1 will collide with Train C1 at destination. Therefore, unless the dispatcher instructs the train crews with correct track authorities, this routing is not feasible.

The Unimpeded Time-Space Diagram shows the hypothetical train movements assuming there is infinite capacity between the nodes:

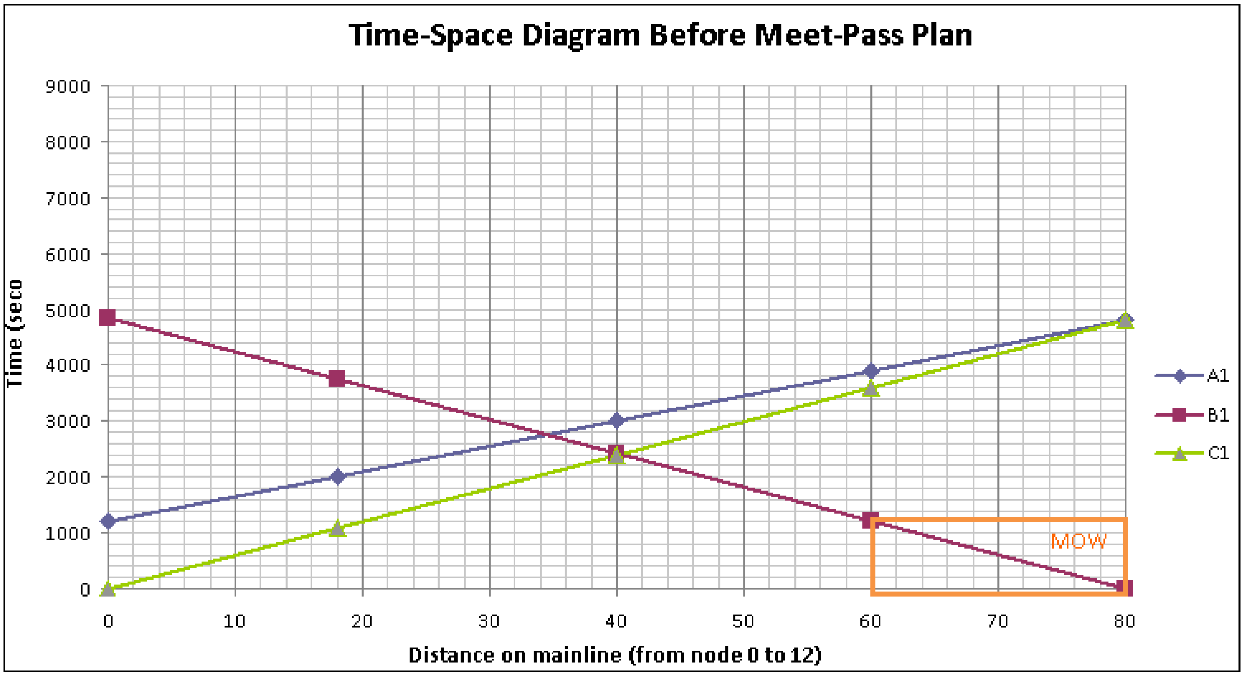

Figure 5: Unimpeded Projections.

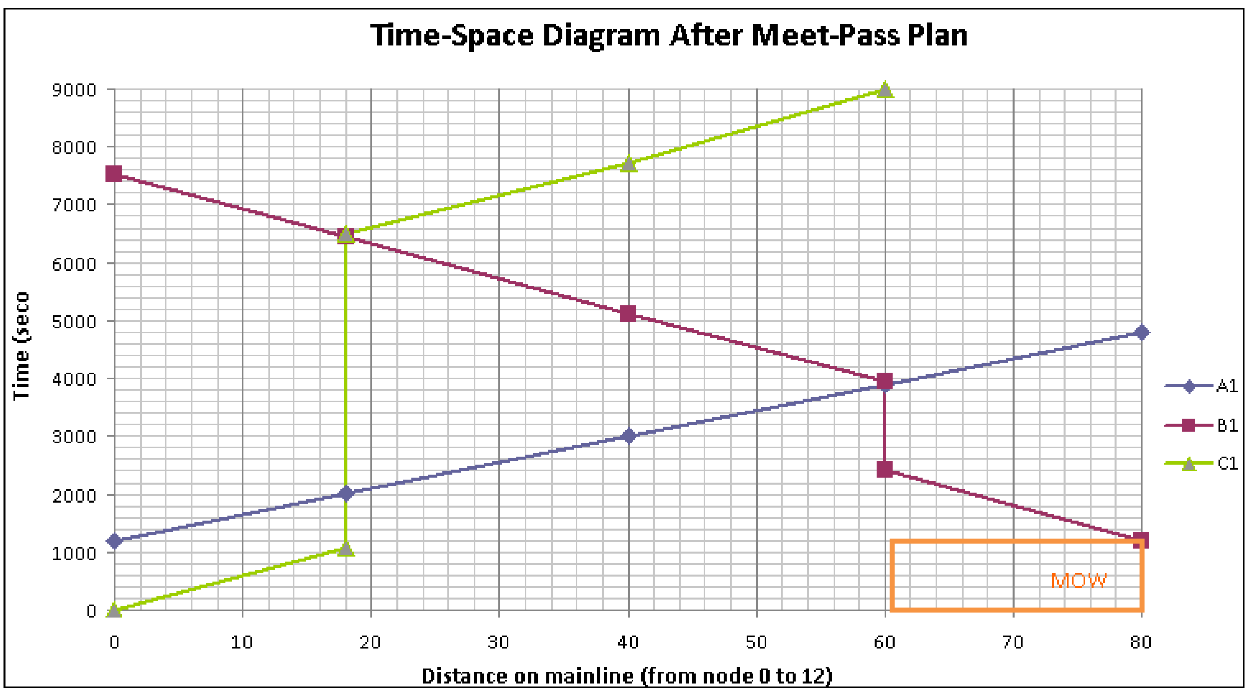

Figure 6: Resolved Plan (not drawn to scale because some trains will travel more distance than others).

Resolution:

A feasible but not necessarily optimal plan can be constructed as follows. According to the data Train A1 is faster, then Train B1 and then Train C1. Because of the MOW window, Train B1 waits until time 20 at node 12, then continues to node 11, takes the siding (10,9) and holds until Train A1 passes by. In the meantime, since we know that Train A1 is faster than Train C1, Train C1 is instructed to take the siding (2,4) until both Train A1 and Train B1 pass by. Train B1's total delay includes the wait for the MOW window. This results in a system with the statistics shown in OUTPUT TOY text file. There is a READ ME for OUTPUT text file which describes how to read this file. Here is the cost analysis for that output:

Table 3: Statistics for the Resolved System (all numbers are in seconds, unless mentioned otherwise)

	Delay Time	Arrive at	Arrive Time	Required SA	SA diff	SA Penalty (> 2 hrs)	Required TWT	TWT diff	TWT penalty (outside 4 hr window)	Unpreferred Track
Train C1	4929.668									0
		node 6	7713.668	3000	-4713.668	0				
		node 12	10113.670	6000	outside planning horizon		7200	outside planning horizon		
Train A1	0									0
		node 6	3000	2400	-600	0				
		node 12	4800	4800	0	0	8700	3900	300	
Train B1	2307.416									0
		node 6	5110.576	-3000	-8110.576	-910.576				
		node 0	7530.744	5400	-2130.744	0	4800	-2730.744	0	

In Table 3, the cells in blue are the total delay incurred at the terminals, the main track or on the sidings, the sum of the deviations from the schedule adherence, the deviations from the terminal want times, and the total time spent on the unpreferred tracks. The amount of time spent on the unpreferred Tracks is 0 in this case, because the toy problem is given for a single-track dispatching territory.

To use the cost figures given the Objective Function section, the total cost for system can be calculated as follows:

Total Cost = (Total Delay * Delay Penalty/Hour) +

(Schedule deviance over 2 hours for SA Trains * Penalty over 2-hour deviance/Hour) +

(Terminal Want Time deviance beyond the 4-hour window * Penalty for TWT/Hour) +

(Unpreferred Track Time * Penalty for Unpreferred Track Utilization/Hour)

= ((4929.668/3600) * 400) + (0 * 600) + ((2307.416/3600) *500) +

((910.576/3600) * 200) +

((300/3600) * 75) +

(0 * 50)

= $925.053

There is no cost impact of Schedule Adherence despite the fact that B1 is late at destination terminal because Train B1 is still hot being less than 2 hours late. Train C1 reaches destination beyond the planning horizon, hence we do not incur any penalty for schedule adherence and terminal want times at that node. Furthermore, Train A1 is 65 minutes early at destination. Based on 4 hours penalty-free window, it incur penalty for 5 minutes outside the window.

The participants should provide a table similar to Table 4 in their reports for all the datasets provided. It could be put in appendix and will not be counted towards the total number of pages requirement.

References

D'Ariano, A., Pacciarelli, D., Pranzo, M., 2007. A branch and bound algorithm for scheduling trains on a railway network. European Journal of Operational Research 183 (2), 643–657

Harrod, S., 2011. Modeling Network Transition Constraints with Hypergraphs. Transportation Science. 45(1) 81-97.

Törnquist, J., Persson, J.A. 2007. N-tracked railway traffic re-scheduling during disturbances. Transportation Research Part B, 41(3). 342-362.

Carey, M., Crawford, I., 2007. Scheduling trains on a network of busy complex stations. Transportation Research Part B, 41(2), 159-178.

Zhou, X., Zhong, M., 2007. Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds. Transportation Research Part B, 41(3), 320-341.

Şahin, G., Ahuja, R. K. and Cunha, C. B., "Integer programming based approaches for the train dispatching problem," 2006.