# Integrated timetabling for section and suburban railway operations: performance objectives and optimality measures

Narayan Rangaraj*, Sudarshan Pulapadi, Swapnesh Subramanian and Madhu N. Belur*
Indian Institute of Technology Bombay
Email*: narayan.rangaraj@iitb.ac.in, belur@iitb.ac.in

## Abstract

This paper encapsulates some of the issues that arise in integrated timetabling: both section rail-traffic and suburban rail traffic. Various performance objectives are considered and this paper elaborates on some of these objectives. In the context of section traffic, we consider the case of mixed rail-traffic, i.e. the case when the resources are shared between passenger train and freight train. Various notions of capacity of a section is discussed. In the context of suburban traffic, we also summarize the relevant aspects of crew-allotment. We describe some recent and ongoing attempts in this area to develop systematic and integrated tools for timetabling, that encompass fixing of schedules, rolling stock utilization and crew planning and also the impact on punctuality and reliability of services that operate on a given timetable. This paper extends a note prepared for a recent Niti Aayog sponsored study, and builds on recent work towards providing technical solutions for suburban operations for Central and Western Railways in Mumbai.

# 1. Some principles of rail-traffic timetabling

The UIC 406 document [Uic04] compiled by the International Union of Railways proposes a framework that is broadly applicable to many of the professional railway operating environments, worldwide. It identifies four considerations in capacity assessment:

- throughput (number of trains),
- heterogeneity (mix of trains),
- stability (robustness) and
- traversal time (service).

Out of these four considerations of capacity assessment, throughput and heterogeneity are directly conflicting (i.e., it is possible to have more number of trains if one reduces the variety and streamline the traffic and conversely, the more the types of trains one wishes to run, the smaller would be the number of trains that one can typically operate). Similarly, traversal time and stability are also directly conflicting - the less the (planned) traversal time, the less room there is for recovery, punctuality and stability of schedules, and vice versa. Similar considerations apply to timetable construction and the performance of traffic on a given part of the network.

The following figure shows this for two types of infrastructure usage:

A. Suburban trains (like metros, dedicated usage by homogenous non-overtaking trains)
B. Mixed traffic (i.e. trains with different running/halt characteristics, typically running on main line sections of Indian Railways)

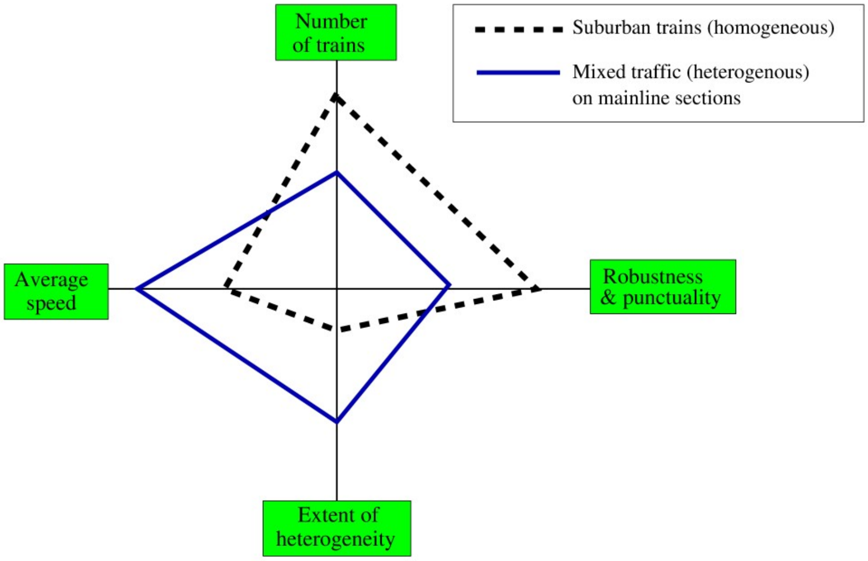

Fig 1.1: The balance of railway capacity (adapted from [Uic04])

The interaction effect between other pairs of performance measures and the combined behaviour is less straightforward to quantify and depict. We summarize these parameters diagrammatically and list three factors on what they depend.

**Traversal time:** Traversal time on a given section depends on

(a) length of section to be traversed
(b) achievable speed of rolling stock (moderated suitably to allow for realistic running)
(c) planned halts
(d) time for accelerating and decelerating to and from halts to allow for overtakes
(e) explicit allowances to account for temporary speed restrictions and traffic hindrances.

An analysis of these factors on important sections of Indian Railways is an ongoing exercise, which is done at different levels - divisional, zonal and at the railway board. This exercise can benefit much from data analysis and regular assessment of operational performance.

**Congestion and Overtaking:** The current method of estimating line congestion is using a variant of standard formulas that are relevant for a single stream of traffic, and are not relevant for mixed traffic where there are significant interactions between different types of trains (speeds, lengths and priorities). In an environment with different types of trains,

- either there is overtaking in which case there are slow movements to allow for this, resulting in some loss of capacity on critical sections,

- or there is no overtaking and then the faster trains are constrained to move at below their maximum speeds, possibly resulting in some loss of potential throughput.

For example, on extremely high-density sections such as Mumbai suburban, the latter strategy is followed and even the high priority trains capable of 130 kmph run at considerably lower speeds and overtaking is completely avoided. The resulting loss of traversal time for these high priority trains over these relatively small sections is considered acceptable because the overall throughput is improved due to the absence of overtaking in the high-density section.

**Junction/Terminal Operations:** A significant aspect of infrastructure that is under-studied is the impact of junction/terminal operations on sectional running. It is a commonly encountered phenomenon that trains (both passenger and freight trains) undergo substantial waiting just short of major junctions. It is worth emphasizing that it is not just platform or running line resources at junctions that are the bottleneck to throughput, but the crossovers and track resources which provide access to various parts of the junction that are often the bottleneck to throughput. The effect is most significant when there are reversals, loco changes and other less-streamlined operations at junctions. This analysis is pursued in more detail in Section 5 below.

The distance-time chart below contains the time-axis as the horizontal axis and the distance along the vertical axis.

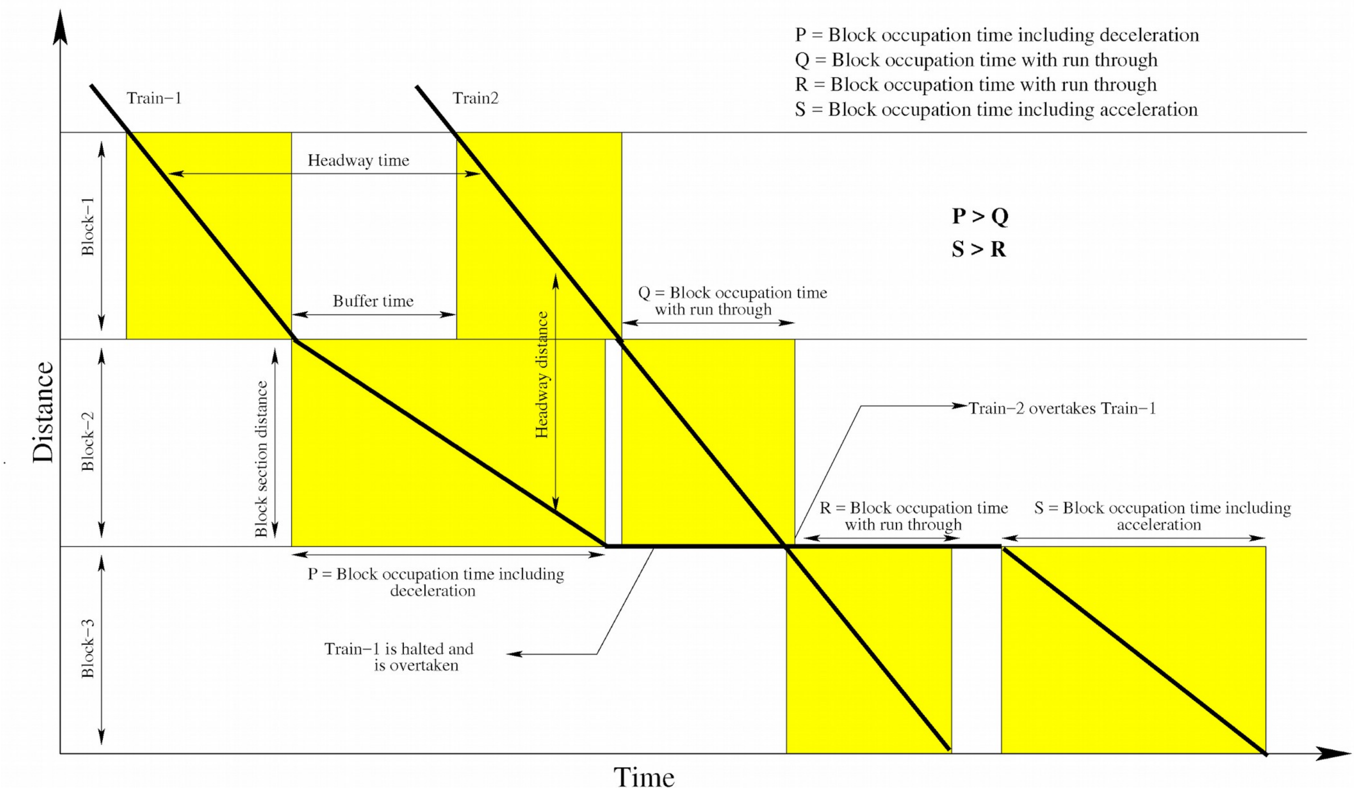

Fig 1.2: Distance-time chart showing headway distance, headway time, and overtake of Train 1 by Train 2

## Good quality paths for passenger trains

With passengers as primary beneficiaries of the railways, below are some notions defining quality of a passenger train path.

- High traversal speed between stations
- Low overtakes by other trains
- Halts only at scheduled stations (and not for just being overtaken by other trains)
- Punctuality of departure

The above points are clearly motivated by the fact that railways is a *service*.

## Good quality paths for freight trains

Most freight trains in India do not run on timetabled paths. But given the commercial importance of freight trains in determining the viability of long distance rail services overall, on many parts of the Indian Railways network, it is worthwhile to examine how freight movements can be streamlined. Rail movement of freight has significant benefits in energy use, environmental impact and safety. A good quality freight path is determined by

1. High traversal speed between stations
2. Low overtakes by other trains and few halts at locations other than where they need to

The above points help achieve more number of freight paths and better delivery of freight service (by quicker delivery), and delivery at a lower cost (by lower inventory, lower usage-time of wagons, lower usage of locos and crew). Timetabling of good freight paths is necessary for specialized facilities such as the Dedicated Freight Corridors, and is relevant even for mixed traffic sections on other parts of the rail network.

## Robustness and punctuality

A railway system is subject to uncertainties and delays caused by malfunctions or deviations from plans. Delays directly caused by disturbances are called primary delays. The delays caused to a train due to primary delays of other trains are called secondary delays. Due to interdependencies in railway system, a large part of delays in congested systems are actually secondary delays.

Some of the major causes for disturbances and primary delays are

1. Planning - while planning the timetable, error could arise, such as the following: dwell time may be planned shorter than required, length of trains not accounted for and acceleration and deceleration parameters not properly accounted for.
2. Infrastructural failures - malfunctioning switches, broken catenary, failing signals, and maintenance works taking longer than the planned times.
3. Human factors - Driver behaviour and response time in manual operations of route setting and other actions add to stochastic primary delay.
4. Weather and environmental conditions - for example, poor visibility in fog leads to increase in braking distance and decrease in acceleration rate. In some rail sections,

cattle run over and animal conflicts are a serious issue and as of now, no effective solution has been found.

The above factors contribute to delays and due to delay propagation, the effects spread in the railway network in both time and space and eventually causes large secondary delays. Some of the major causes of secondary delays are

1. High capacity utilization and thus smaller headways could lead to speeds that are lower than timetabled ones.
2. When a train reaches a terminal station with a delay, the subsequent train in the rake link could be delayed, depending on the time provided.
3. When a train reaches late, the planned crew at the interchange point also gets late and can not serve their duties in full, thus leading to delay of subsequent train.
4. Late running trains (or early trains) occupying resources that trains that are on time are unable to use.

Train data usually related to delays include both primary and secondary delays. From robustness considerations, a good timetable is one in which the primary delays get absorbed quickly and do not impact other trains and cause secondary delays.

## Criteria for a “quality timetable” for mixed rail traffic scenario

The above notions of quality of passenger trains and freight trains are to be kept in mind when defining the quality of a <u>timetable</u>. These criteria are to ensure higher throughput (in terms of passenger trains and freight trains), better quality paths for passenger and freight trains, and thus better usage of rail-infrastructure.

A. Low traversal times (time spent in the system)
B. High running speeds of trains
C. Minimizing overtakes between trains
D. Punctuality of <u>departure</u> at stations with scheduled halts
E. Punctuality of <u>arrival</u> at stations with scheduled halts
F. Resilience/robustness of the timetable due to unanticipated disruptions

Point E above about punctuality of arrival at a station is important to ensure that trains arrive neither late nor **<u>earlier</u>** than scheduled time: arriving **<u>earlier</u>** uses up valuable infrastructural resources - this detrimental effect of arriving too early appears to be underestimated. On the other hand, some allowances are required and inevitable to ensure Point F above and thus there is a trade-off between points E and F.

The above points are analyzed in more detail by considering the following quantities:

**Allowances:** Allowances are extra time values given within a timetable for each train at regular distances (or at the end): this is given to help maintain punctuality of operation of a given timetable.

**Planned overtakes:** Of particular significance (and to be avoided) are the overtakes between trains of a similar type. Overtakes can also be derived from working timetable by analysis of arrival and departure times at a station. Note that overtakes can happen only at a station.

**Bottleneck resource: definition and capacity aspects:** For our purposes, a bottleneck resource is one whose improvement will result in the improvement of the entire section as a whole. There are different ways in which one could identify a bottleneck resource.

- Section with maximum utilization
- Section with minimum speed per hour (of freight trains, but also of passenger trains)
- After identifying the section definition and signaling regime, the block section that takes longest to traverse

**Train composition:** In order to maximize the benefits of paths, on bottleneck sections, as a general policy, trains of high capacity should run.

**Speeds, halts and overtakes:** In a timetable, one can calculate the number of overtakes within each type of train (i.e. slow-passenger, express trains, super-fast express trains, etc), and the average speeds of each type within a section. In particular, overtakes between trains of the same type are undesirable since this brings down the throughput. Furthermore, scheduled overtakes between trains of the *same* priority are avoidable. While these overtakes are ideally minimized, if a section is identified as a “bottleneck section”, it is crucial to ***eliminate*** overtakes ***completely*** on that section: this helps improve overall throughput.

# 2. Capacity utilization of sections handling mixed traffic

Capacity is often specified as the ‘‘maximum number” of trains per day a section can accommodate, and a typical utilization or “throughput” is compared with the capacity to obtain “efficiency” of the timetabling process or implementation/operation. There are many different definitions of capacity and capacity utilization, some of which are listed below. The third among these (2C below), leads to the familiar “Scott’s formula” which is still used to some extent in the Indian Railways context.

## 2A: Mixed-traffic-ideal-grouping notion of capacity

With homogenous traffic, capacity is the number of trains that can pass through a section in a given unit of time: here per day. The reality in *mixed* traffic is much more complex, as the operation has to consider a specific number of trains of different characteristics and some specific (or at least typical) sequence of these trains on the given section. In such a situation, given sufficient traffic, if there is no overtaking, faster trains are forced to travel at lower than their maximum or rated speeds OR slower trains will be further slowed down as they decelerate, stop, get-overtaken, and then accelerate in order to give way to faster (usually higher priority) trains. In either case, there is a significant loss of capacity.

Earlier studies by Indian Railways have highlighted the impact of speed differential on section capacity. The note [Ran-18] provides some initial basis for objective and meaningful computations of capacity in such situations. This has a significant impact on timetabling.

**Total capacity utilization under the situation of <u>typical-grouping of trains</u>:**

This capacity is calculated by considering the time loss due to overtakes (acceleration, deceleration) for a typical sequence of train running on the section.

**Total capacity utilization under the situation of ideal-grouping-of-trains:**
This capacity is calculated by considering the full grouping of trains (no overtakes) but with the desired mix (numbers) of each type of train.

**Brief explanation of capacity calculation:**

- Time to travel the bottleneck block (longest block) in a section is calculated.
- Actual time utilized is calculated by sum of travel time by trains, prior and later headway for each train and time for overtakes (acceleration time, deceleration time and time to cross the loop ).
- Available time is 840 minutes after excluding 240 minutes (i.e. 4 hours) for maintenance and say 70% efficiency from 1440 minutes (the number of minutes per day).
- Capacity utilization is the ratio of actual time utilized to available time.

A low value of capacity utilization here would suggest trying a proper grouping of trains, thus calling for more careful timetabling, rather than automatically justifying more investment in infrastructure.

### 2B: Distance-time-chart-occupied notion of capacity utilization

A second way to define capacity is the fraction of area “occupied” by trains with respect to the total area in the distance-time chart. It is essential to include safety related headways into the notion of “occupied”. This definition allows a 100% capacity utilization by trains as long as they have the same running characteristics, independent of their running speed. One can introduce the 70% efficiency and maintenance block margins here too. Though this definition has its merits and has been studied in the literature, we do not pursue this definition in this report. A low value of capacity utilization obtained by this definition would also suggest some leeway to be achieved through proper grouping of the trains in the mixed traffic scenario.

### 2C: Bottleneck section: slowest-train-based notion of capacity

A third notion of capacity would be the “throughput”: here the bottleneck section is identified and the number of trains with slowest speed is calculated and the actual number is compared with respect to this number. Conventionally, line capacity has been calculated in the Indian Railways context using the so-called “Scott’s formula” which involves the number of trains that can pass a section: the deciding factor being the slowest train and the bottleneck section. Headway constraint is also taken into account, and a factor of about 70% to 85% is included to ensure resilience. This definition of capacity is widely used in Indian Railways and several references in Indian Railways documents contain an elaboration: see [Ope17], for example.

This notion suffers from the drawback that since quite a few trains run at speeds faster than the slowest train, we get an exaggerated capacity utilization percentage (often much higher than 100%); the obtained value of the capacity utilization is too sensitive to the fraction of the slowest trains in the entire day’s train composition. We note that this definition is meaningful to some extent, in homogeneous traffic conditions, such as suburban sections or freight only sections.

A low value of capacity as per this definition would call for splitting of the longest block and/or increase of the speed of the slowest train. Investment in infrastructure to reduce the traversal time over this bottleneck would help in capacity improvement. A low percentage capacity utilization would mean much heterogeneity and improper grouping of trains.

### 2D: Bottleneck section: fastest-train-based notion of capacity

One can consider the fastest train for the purpose of capacity calculation: we then have actual percent utilization (in the presence of mixed traffic) as lower than 100%. If the number of the fastest trains is much lower than the rest, then this definition would give a very low value of capacity utilization.

A low value of capacity here would mean that the bottleneck section needs more infrastructure upgradation and a low utilization percentage would mean much heterogeneity and improper grouping of trains.

It must be noted that “throughput”, as a quantitative measure of how much a given capacity is utilized, and “congestion” is a standard term across many areas involving flows over networks: air-traffic, internet data-traffic (see [Odo87], [Low04] and references within for some examples). A *qualitative measure* of capacity utilization is as important too: this ensures that customers are satisfied by the Quality of Service (here, the customers are railway passengers and freight traffic beneficiaries.)

The following quantities play a key role in the computation of capacity.

#### Headway Time (minutes)

Safe operating headways are applicable between different categories of trains. The prior headway reflects the importance in maintaining punctuality of that category of train and the post headway reflects the requirement of the lower priority category of train to be able to follow the leading train at sufficient speed. We note that the prior headway for the higher categories and the post headway for the lower categories of trains that are important.

#### Capacity utilization of some sections

Using typical parameters and some representative numbers for the traffic on various sections and typical sequences that are observed, we can compute capacity utilization. We can also quantify the extent of improvement of this with resequencing in order to streamline traffic.

# 3. Analysis of factors that affect congestion

This section discusses various parameters that affect a timetable and the congestion caused due to primary and secondary delays.

## Maintenance block considerations

1. As per current practices, in order to ensure safe running of trains, it is essential to have maintenance blocks of duration between 2-4 hours, contiguous in time (usually in day time hours) and for an entire section at a time. This has a significant impact on available

capacity. In future, with preventive maintenance and condition based maintenance with sensors and other inputs, this may need revision. In advanced railway systems of the world, there are devices that are deployed to collect and monitor the health of the track and other infrastructural components in a faster way and the data is analyzed in real-time at a railway-data-center for fault detection and diagnosis. While expertise for deployment and monitoring requires to be developed, this is the systematic and future way, and moreover, this is not a large infrastructural investment but is quite a viable tool. The monitoring aspects of maintenance blocks is avoided and capacity can thus be improved.

2. On congested sections, there is a conflict between the requirements of maintenance and traffic goals: an improper planning leads to congestion: though the congestion is local and temporary, it typically takes considerable time to resume normal operations and recover the lost time.
3. After any substantial maintenance work, there are speed restrictions on the track due to which train travel time increases (average speed decreases). This is part of the allowances built into the operating timetable on any section. Generally speaking, these are proportional to the length of track that has to be operated.

## Slacks and allowances

In order to obtain good punctuality of railway operations, small disturbances and delays need to be accommodated during the operation: a timetable needs to absorb such durations of time. This is usually explicitly listed in a working timetable as ER/OR (recovery time values). Since the calculation sometimes tends to be ad hoc, and can have a very negative impact on both capacity and punctuality, we dwell on the calculation procedure and then on the consequence improper allowance in this subsection. The allowance can be calculated in two ways:

(1) by scheduling trains at speeds lower than technically achievable - we refer to this as *slack* and

(2) by keeping scheduled running times longer than the technically minimum running time - this difference is known as *allowance* or running time supplement.

Slacks are planned keeping in mind driver behaviour and equipment performance in safe regimes.

For allowances, there are two steps that need to be followed to have punctuality in the operations of a timetable.

**1. Determination of total allocation to be allocated to a train throughout its journey**

Total allowance should neither be too high (as this leads to longer scheduled travel time, which in turn consumes the capacity of the system) nor should it be too little (as this will not help achieve punctuality in the operation due to the presence of disturbances and daily exigencies). Generally, allowances are given in terms of percentage of the minimum travel time of the train. Internationally, the total allowance varies between 3% to about 10% of the running time, depending on the train and route characteristics. This amount is 15% or more for some trains in Indian Railways, which is on the higher side.

**2. Optimal distribution of the total allowance to all stations throughout the journey**

Once the total allowance to be allocated is decided, the next step is to distribute it among all stations in a rational way. One can allocate the allowance at the end of the section or alternatively distribute it at intermediate stations. There are merits and demerits to both as summarized below.

Though the chance that the allowance will be used effectively if all the slack is given to downstream stations is high, the train often gets delayed at earlier stations also and this causes knock-on (i.e. ripple effect) delay to other trains and thus affect the punctuality of the other trains. On the contrary, if more allowance is given to earlier stations and if there are no early disturbances, then the early allowances are lost and cannot be used. Since the delay reduction at a station not only reduces the delay of that particular train but it also reduces the knock on delays for subsequent trains, so instead of evaluating the delay at last station only, the sum of delay at each station should me minimized. Systematic methods need to be developed to rationally apportion slacks and allowances.

## Grouping of trains based on speeds

Grouping of trains means to schedule trains of similar running characteristics in a bunch rather than interlace them. Grouping of trains leads to minimum overtakes and thus less traversal time. Grouping thus also leads to better capacity utilization, this means one can enable the running of more trains for a given infrastructure. Grouping however sometimes cannot be done to the desired level due to requirement of passenger convenience of timings of scheduled trains. Notwithstanding this requirement, the principle of grouping should be followed to the extent possible and should be a major consideration at the time of preparing timetables, especially for newer varieties of rolling stock.

# 4. Junction congestion analysis

The analysis of congestion at a junction is significantly different from that of a section on a network. At a junction, various combinations of simultaneous or near-simultaneous movements are possible and resources have to be shared more carefully. The relationship between capacity (numbers of trains, even if we define a mix), delays encountered in traversal times and resource utilization is not as clear as it is for sections. There are currently no well-defined frameworks or planning tools to assess junction resources in a systematic manner. Some of our studies have been reported in [Mig17-2, Ran02, Sal17] and this still remains a topic of research worldwide.

In order to quantify how much a entry/exit layout at a junction can affect the “throughput” of a junction: one can use a so-called resource-to-resource hindrance analysis. This method is not as applicable to a *section* analysis since in a section, near simultaneous movements are not allowed (except in opposite directions, or during an overtake).

The case of Allahabad junction (ALD) with 19 lines (both up and down) and 10 platforms is a good example of how hindrances on a train movement due to movements of other trains can use-up a lot of time and resource. When a train arrives on a line, it not only causes hindrances to the platform or line on which it is halting/passing but also to other lines as the lines need to share various common linkages during the process of entering/exiting the station. A hindrance matrix is constructed to tabulate these hindrances. (Details of the approach can be found in [Mig17-1, Mig17-2].) Analysis of the hindrance matrix along with the timetable of passenger trains at Allahabad junction helps identify which platforms are under-utilized, which ones are unavailable due to cross-movements into/from other platforms: this can help in decisions involving redesigning of junction layouts.

Allocation of platforms to passenger trains in an optimal way can also help in decreasing the hindrances. One can allocate platforms to passenger trains (at least to some trains) in such a way that the train’s entry/exit-movements produce very little hindrance to the other trains entering/exiting at that time.

A junction entry/exit line layout can be termed perfect if for each platform, we have either availability (blue) or the platform is occupied (green): the hindrance amounts are ideally zero. The hindrance amount indicates the extent of non-availability of a platform due to a cross-movement of another train.

We propose the following for planning freight trains at a junction.

1. Plan a time of day dependent variable freight halt so that one can get a more realistic picture of the freight trains waiting/movements at large junctions.
2. Backtrack the freight trains and start them from their source station such that they experience small amounts of hindrance due to other train movements while entering/exiting congested junction.

An analysis of a junction using this approach, followed by a routine sensitivity analysis, can help identify the bottleneck links at the entry/exit of a junction: it would be incorrect to term the shortage of platforms as the bottleneck since, as is visible from the above plot, most platforms are occupied by less than 25% of the day.

A careful planning of short linking lines (to cause small hindrance to cross-movements) combined with a systematic platform allocation can reduce the hindrances in the hindrance matrix. This approach has been demonstrated above for Allahabad junction and for Kanpur area in [Mig17-1]. This approach is a possible line of analysis for understanding a complex operational issue, acknowledged across the world as a complex one, and that is a matter of ongoing research.

More recently, a simulation based approach was used to understand movements in the Ahmedabad junction area in detail. This included all relevant primary movements of passenger and freight trains as well as auxiliary movements of (a) loco attach/detach for terminating/originating trains (b) loco movements to and from loco sheds where there is a change of traction, and (c) movement of rakes to and from the maintenance lines. Three layouts were analyzed, (a) the existing layout, (b) a layout proposed by the division that seeks to create ***through*** lines in both up and down directions and (c) the possibility of additional crossovers to increase flexibility. Detailed simulation on these infrastructure options provides useful insight that allows planners to decide on options and to sequence the movements in a near-optimal manner.

# 5. Typical performance objectives: guidelines

The guidelines below arise from a combination of tools and analyses performed from the data that was used railway related studies conducted by our group. This area deserves a much more extensive and continuously updated efforts on the part of Indian Railways, as conditions are bound to be evolve in the years to come.

**Capacity-utilization:** Capacity utilization measures are used as an indicator of congestion and high values are used to justify additional investments and possibly as a justification to provide for more traffic allowances in timetabling. While the general principle is acceptable, the capacity measure itself needs significant refinement and consensus in use before any valid claim can be made in mixed traffic sections.

We suggest a simple extension that is formula based, to start with, but which considers both the mix of traffic on a section and to some extent, the sequence of traffic on a section. Further, since congestion is a direct consequence of both:

- timetabling (grouping, allowances and other scheduling aspects)
- inadequacy of infrastructure (additional main/loop lines, signalling upgradation needs,

etc),

decisions about additional investments must be backed by a simulation and comparison of the simulated timetable before and after adding infrastructure: IIT Bombay's mixed-rail traffic simulator is one of the available tools, another tool, for example, is Satsang by CRIS.

**Bottleneck-scheduling:** A principle in bottleneck scheduling in multi-resource environments is that the maximal throughput strategy on the bottleneck resource should drive the schedule on the rest of the network. If a section appears to be a bottleneck using any of the above measures, then the traffic on this section should be streamlined, with as few overtakes as possible, and also as ideal grouping as possible, so as to achieve maximum throughput and overall traversal time performance.

**Junction-analysis:** The impact of junction movements is very significant and causes significant cascading impacts on sectional running as well. This area of analysis in railway operations is a very challenging one worldwide and proper tools and techniques need to be developed to analyse this. We have suggested some beginnings in this paper.

**Slack-and-allowance-distribution:** Divisional measures of punctuality should be reworked to have a more continuous unit of measurement (rather than a slab based measure) and should be based on the resources available in each part. The current practice of loading all allowances at the end of a section before interchange is detrimental to punctuality of operation and should be re-evaluated.

It is recommended that only reasonable amounts of allowance is provided for section-congestion, and the provision be focussed on easing section-congestion and not at easing junction-congestion. For guarding against junction congestion, given the occupancy pattern of platforms at junctions, it appears that additional halt times at junctions be explored. These options should be re-evaluated from time to time as network performance measures are monitored.

We now discuss timetabling and related planning decisions on a specific part of Indian Railways, namely suburban networks, in detail.

# 6. Rolling stock management: standardization and impact on punctuality and rake utilization

The quantity of rolling stock required for providing adequate levels of service should be carefully determined, as rolling stock forms a significant cost component in a railway system. According to Raghuram efficient rake links will lead to effective utilization of coaches [Rag86]. Efficient rake links will lead to effective utilization of rolling stock by minimizing the number of rakes required for running a set of services as per the timetable. One factor that hampers efficient rake linking is the difference in composition of rakes. While customized service is welcome, this also puts some constraints on asset utilization and flexibility in operation. Use of quantitative modeling in rolling stock management is necessary to understand the extent of improvement in utilization prior to standardization of rakes. In addition there is a need to understand the punctuality aspect of service when there is a possibility of trip delays in the system.

A direct benefit of standardization is the possibility of needing lesser number of rakes for running a given set of services. If the rakes are standardized any rake that has arrived and completed maintenance can be used for a subsequent trip. The pooling effect could result in the use of lesser number of rakes for a particular set of trips. There can be other incidental benefits like uniformity in maintenance practices leading to lesser maintenance times. There have been some instances where Indian Railways has standardized the rakes.

Many complex problems in railway systems can be modeled as combinatorial optimization problems [Cap07]. Network flow modelling and vertex colouring approaches are used to analyze the aspect of improvement in rolling stock utilization. The objective is to minimize the total number of rakes required for managing all services as per the given time table. Vertex colouring problem is a popular and well researched combinatorial optimization problem. Marx has discussed the scope of using graph colouring models in solving scheduling problems [Mar04]. In the vertex colouring approach, each node in the graph represents a scheduled trip to be run as per the time table which has a start time and finish time associated with it. Any two nodes have a link (arc) if there is an overlap of time intervals between the trips, i.e. there will be an arc from a node i to node j if it is not possible to start a trip j after completion of trip i. It means that the start time of trip j is earlier than the finish time of trip i. It is modelled using an integer programming formulation. In the graph colouring model the nodes assigned to one colour corresponds to one rake. The priority in use of the rakes for these trips (nodes) is dictated by the precedence relation of trip timings. Once the colours (rakes) for the trips are obtained by solving the model, a small computation (arranging the trips of the same colour in the ascending order of trip start timings) is required to get the rake links.

The same problem is modelled as a Minimum Cost Flow problem in a network [Ran06]. The directed network is represented as a Graph, G = (N,A), which is formed by a set N of nodes and a set A of directed arcs. Each node represents a trip and a directed arc is created between the nodes if there is the feasibility of such a link. Each arc (i, j) in A is associated with a cost $c_{ij}$, which is the cost per unit of flow in that corresponding arc. The problem can be modeled as a Minimum Cost Flow problem by suitable modification of the sequential graph of trips. In the graph (shown in figure:6.1 corresponding to trips shown in Table:.1), each node representing the trips (eg. node A) is split into two nodes and arc is created between the nodes. In addition a source node is created and it is connected to the nodes representing each trip. Similarly each of the newly created nodes (eg. A') are connected to the sink node. An arc is also created between source and sink nodes. A directed path in the network corresponds to a feasible sequence of trips. For example there is a directed arc between A and C, since trip C can be started after completion of trip A. The source node is given a supply of N, where N is the number of rakes available and demand for this node is 0. For sink the supply is 0 and demand is given as N. For all other nodes the demand and supply values are given as zero. The flow capacities and costs of the arcs are chosen so that the desired result is obtained (ie. minimum rakes are used).

The objective is to minimize the total cost of flow through the network subject to flow balance constraints of each node and flow capacities of each arc. The variable is the number of units (rakes) that flow from node i to node j through arc (i,j). In the Minimum Cost Flow network model a flow of one unit from the source node to a node through a network supply edge indicates the use of a new rake. Unit flow from any node to the sink node indicates that for no more further trips that rake is used. The flow of rakes between trips (use of the same rake for two trips) is given by the flow in direct link edges having value one. This model directly gives the rake links. The Minimum Cost Flow model gives some flexibility in extending the analysis by considering the trade-off between costs and benefits of standardization.

Table 6.1: Trip details

| Trip Name | Start time | Finish time |
|---|---|---|
| A | Day 1, 6:00 | Day 2, 16:00 |
| B | Day 1, 11:00 | Day 3, 14:00 |
| C | Day 2, 19:00 | Day 5, 7:00 |
| D | Day 3, 11:00 | Day 6, 18:00 |
| E | Day 5, 6:00 | Day 7, 20:30 |

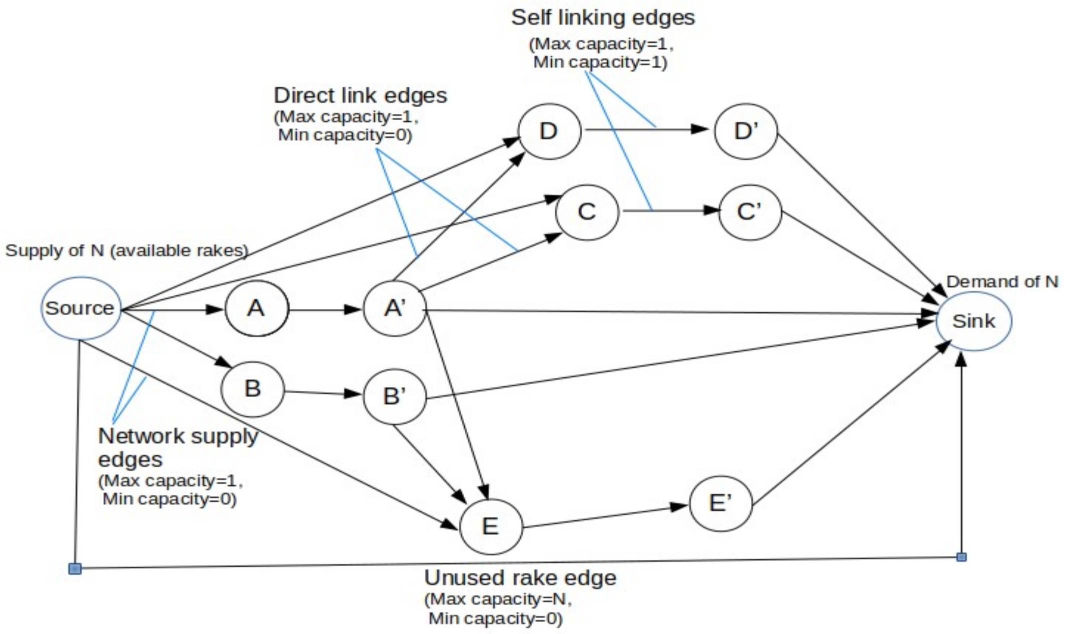

Fig 6.1: Minimum Cost Flow Graph

It is found that in some subsets of services there is a possibility of improvement in rake utilization by standardization. Even though both models gives the same solution, the Minimum Cost Flow network model is computationally more efficient than the vertex colouring model which is reflected in the computational time while running the models. For large instances, the run time of the graph colouring model is in days where as the network flow model provides the solution in seconds.In practice there are different sources of uncertainty in railway operations like run time delays during a trip. The analysis was extended to include these uncertainties. A simulation model was developed to analyze the impact of random trip delays on on-time departure of trips and delay propagation. It is found that in the case of standardized subsets, delay propagation stays in control even for fairly large values of delays and the system is able to recover fast from the effect of big-disruptions in between.

# 7. Suburban railway planning

Suburban rail services in urban or regional geographies generally comprise of homogeneous services (e.g. common speeds and rake compositions) which are run at high density, so high throughput is desired.  The stages of planning required in suburban timetabling include:

- **Line planning:**  This includes planning of number of services required between stations. Optimization in this stage depends on the demand requirements. In consideration to peak demands higher frequency of services is required to maximize the flow of traffic.
- **Timetabling of services:** With the given requirement from line planning, timetabling of services involves generation of schedule for services between the stations. Here, the optimality adds to the robustness of the system.
- **Rake linking (vehicle planning):** After the generation of timetable, rake planning involves allocation of rakes to the services. Optimizing rake linking affects immensely on the network because of lesser number of rakes running at a given time improves robustness of the system. So for the same number services there would be lesser delays. Indirect effects of optimizing rake linkages is more frequent maintenance schedules to rakes.
- **Crew scheduling:** crew planning involves in rostering of crew schedule to operate the services. Improving crew schedules affect directly to the cost of operation to railways. Robustness of crew schedules depend highly on the robustness of the timetable and hence the number of crew required on hold (back up crew).

In the following sections we describe the above stages of planning and different tools developed to generate good results.  Details are available in [Jai19] and [Kas19].

## Timetabling and Rake linking:

Timetabling in India is presently done manually with some computer based visualization and decision support. The planner schedules trains based on track availability and historical demand patterns. This is an iterative procedure which starts by modifying the already existing timetable. The existing approach completely ignores any sort of optimization that one might use while designing such timetables. So it is difficult to manage a de-novo construction of timetable, also it becomes difficult to maintain optimality during maintenance breaks. So this calls for an optimizer tool to generate timetable.

Our approach uses a constraint representation and then solution of a MILP (Mixed-integer linear programing). The following are the inputs the program takes to generate timetables:

- **Infrastructural parameters** like stations, path between stations and traversal time.
- **Services information** like types of services and number of services (desired frequency between stations)

The tool generates a constraint model for the services schedules using the inputs. Each of the variables are linked to other events via constraints which govern the limits of difference between events. The constraints used are:

- **Headway constraint** is a safety constraint which limits two services to be operated too close to each other.
- **Dwell time constraint** limits the stop time at a station.
- **Turn-around constraint** is the halt time at terminal stations where the rake changes service.
- **Traversal constraints** are assumed to be hard constraints, which describe the traversal time between stations.
- **Platform constraints** are added, which keep in consideration about the infrastructure design.
- **Rake constraint** limits the number of available rakes, which in addition helps in rake linking.
- **Frequency constraints** puts an upper/lower bound between services on any O-D pair, ensuring uniform distribution of services.

The constraint model satisfies the strict requirements on services during peak hours, but to generate a valid day long timetable, additional requirements are needed like stabling points of the rakes and off peak demands of services. So in a given day there are two peak periods, morning peak (8:00 am to 11 am) and evening peak (5:00 pm to 8:00 pm). The above procedure is used to generate the peak period timetable. For the off peak timetable, the same algorithm is run with some additional considerations, but liberalising many of the peak duration constraints. These off peak schedules include early morning off peak (before 8:00 am), noon off peak (11am to 5 pm) and late night (8 pm onwards).

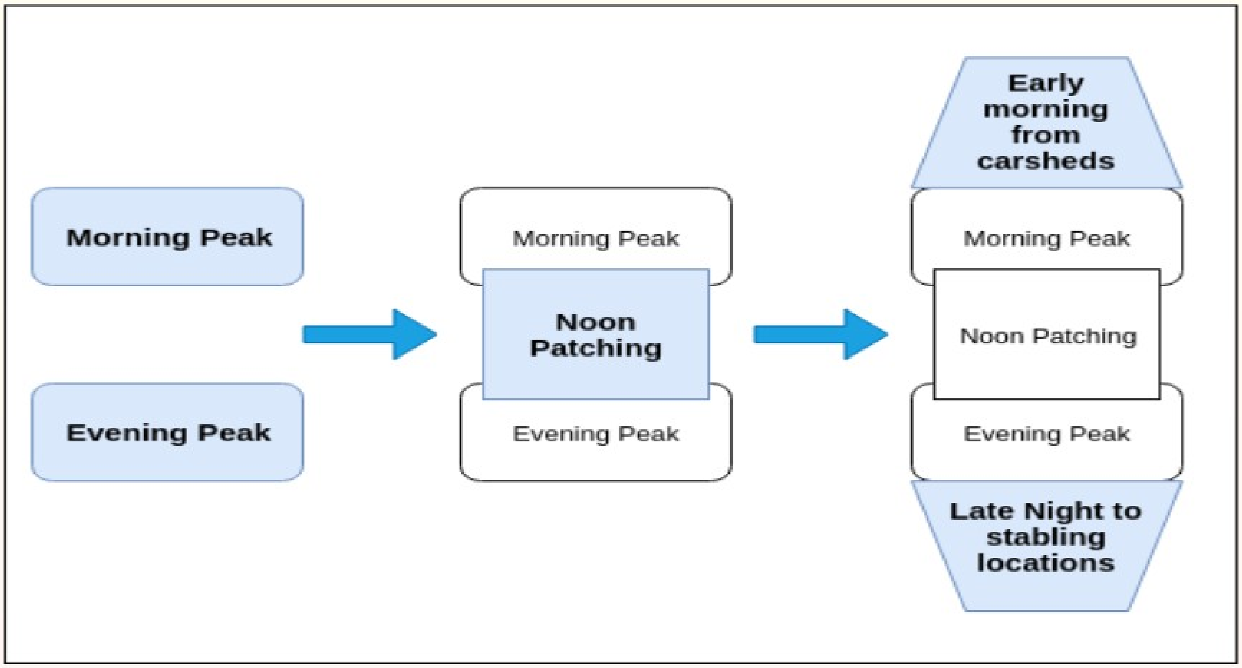

Fig 7.1: Steps to construct complete day timetable, from [Jai19]

The final output of the program is arrival and departure event values and also the rake cycles which define the sequence of services a rake has to perform in the timetable.

**Crew scheduling:** Crew scheduling for suburban services is a detailed planning activity which in practice takes 2-3 months to plan based on the timetables. It involves allocating various services from the given timetable into sets of duties that each crew has to perform on a given day. This scheduling depends on various factors like Timetable, Maintenance breaks etc which are being modified frequently. This creates a need for semi-automatic optimized tool for crew allotment.

The problem in crew scheduling optimization involves finding the correct services to group in a set for a day's work of a crew member. To create an efficient strategy for crew allotment, the overall problem has been decomposed into the following 2 stages:

- **Set generation:** All the services are grouped into set of duties which are less than 8 hours. Each of these sets are the working shift of individual crew. The objective is to minimize the number of sets generated, which directly minimizes the number of crew required to run all services and also the sets generated must have an equal share of workload i.e, sets should be fairly uniform. There are three types of sets:
    - **Day Set :** These sets start after 6:00 am and cover a max 8 hours of duty.
    - **Night set :** These sets start after 10:00 pm and end early in the morning and comprise of fewer services than day sets
    - **Halting Set :** These sets occur in pairs, which are evening duty and morning duty, with a minimum of 5 hours rest in between
    - In addition, there are **Shunting Duties,** which do not contain any services but require crew to take rakes to/from stabling depot (yard or car shed)

- **Set linking:** This organizes the sets in sequence in specific order, which will then be performed by a crew. This defines the duty of crew over multiple days, including the rest hours between the sets. The objective here is to maximize working hours in a given period and minimize break periods in between.

These set definition and set linking decisions have to follow HOER (Hours of Employment and Period of Rest) rules, which describe the constraints for formulation of these sets. In addition to HOER, there are considerations that arise out of field expertise, operational knowledge and practicality of schedule preparation. These result in more than 30 constraints, and a few of these are listed below:

- **Work duration :** Each set must be less than 8 hours long and total working hours in halting set must be less than 14 hours with a 5 hour break in between.
- **Breaks :** A minimum of 10 minute break must be given between each service in a set and a 30 minute break must be given during meal breaks.
- **Weekly work duration :** Total hours worked in a week must be less than 52 hours.
- **Rest periods :** A minimum of 12 hours must be given after each set and 30 hour after night sets.
- The service added to the set must originate from the station where the previous service in the set ended or from a nearby station, to where the crew can travel as passenger (TAP).
- The maximum allowable number of services in a set is 5, preferably a maximum of 4.
- In set linking, night set must not be linked in succession to another night set and similarly for halting sets must not be linked in succession to another halting set.
- Working set must include time to take rakes to/from yard/car shed.
- The limit on the resource available puts a limit to the number of halting sets possible per resting station.
- Night set must not be given more than 2 services.

The above problem has been solved using a flexible and efficient heuristic - using python scripts. This heuristic uses a time weighted probabilistic function to create multiple allocation schemes and a workload balancing function is used to further improve the results. This is an iterative approach of creating work duties, a metaheuristic that is largely greedy initially with a self-correcting mechanism. Creation of large number of allocation schemes, all of which have the constraints enforced, gives us a large subspace of possible solutions with a high probability of finding a good quality solution satisfying multiple objectives in a very large space of possible allocations.

To generate a set the algorithm first picks a service from a pool of services available and starts generating a set. The algorithm checks whether the set can be first used as a halting set, then checks for night set and finally checks for day sets. While generating the sets, the algorithm checks for next services that can be added from the pool of services. This selection is done by a time weighted probabilistic function. The algorithm checks if the crew can continue with same rake, if not it looks for other services originating from present station. After generation of one set, the process continues until all services are allotted after which a load balancing algorithm is used called shuffle and merge, which rearranges the services among the different sets.

For set linking, the algorithm initializes a sequence of sets randomly. Every window of 7 days is checked, to verify that the count of total working hours fails to lie inside a specified upper and lower bound. The upper bound is 52 hours and the lower bound is initialized with a low value, which is incremented as the iterations increase. For violations reported by the checker function, the linking is split so that the remaining links satisfies the constraints. The sets that are removed are returned to the pool of non-linked sets. Further, the algorithm greedily adds non-linked sets to the link to improve working hours. These steps are continued until every part of link satisfies the constraints. This algorithm gives near optimal results in acceptable computational time.

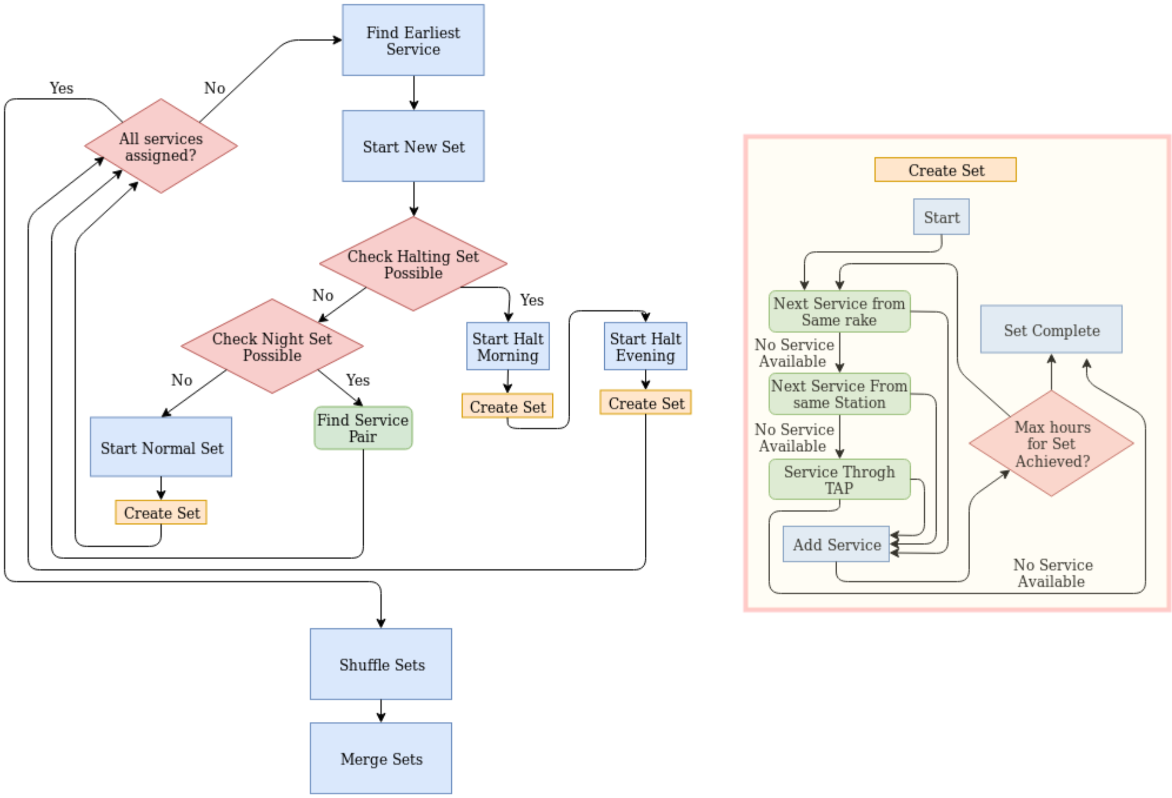

Fig 7.2: Flow chart of set generation, from [Kas19]

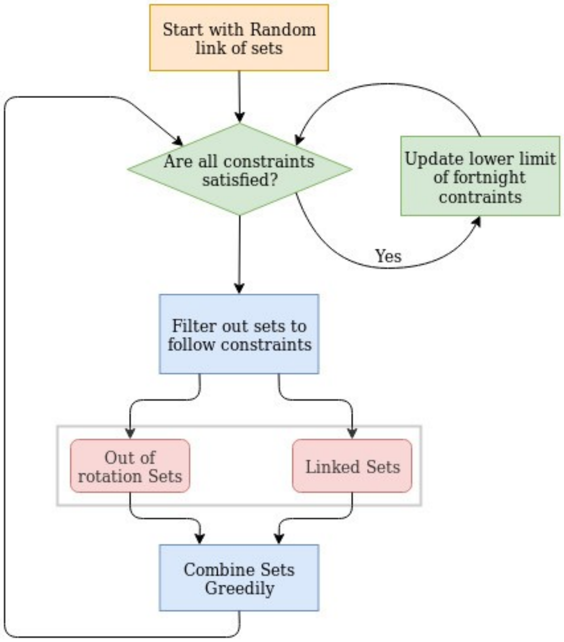

Fig 7.3: Flow chart of set linking [from Kas19]

# 8. Conclusions

Timetables serve as performance targets for train operation and are constructed carefully and with a lot of planning. With the increasing size and number of planning units of Indian Railways, it is essential that this task is done in a well thought-out and planned manner, with as much data-based input and evidence-based decision making as possible, rather than only the subjective inputs of experienced timetablers. For example, the platform provided by Satsang, the timetabling tool of CRIS (Center for Railway Information Systems) is a sound basis for timetabling and should be used more effectively. It appears that this is not currently being done for reasons that are unclear.

Worldwide, railway organizations make use of ongoing developments in planning technology ranging from optimization and operations research to machine learning and data analysis to improve operations. This is true for main line operations, including freight services, as well as niche but important areas such as suburban rail services and metro services. Indian Railways should be the leader in this area, given the complexity and volume of services that it offers its customers and the role it has to play in the socio-economic development of this part of the world.

In India there is high demand for rail traffic and there are severe capacity constraints necessitating conflicting requirements of higher capacity utilization and acceptable level of robustness against delays. Based on preliminary analysis, it is found that the standardization of rakes can result in significant improvement in rake rake utilization, punctuality of services and robustness of operations. The techniques described in the paper can facilitate better decision making in this regard.

In the context of suburban railway planning use of optimization techniques can aid in the activities of timetabling, rake linking and subsequent activities like crew scheduling. The system LinTim [Lin], developed in Germany is an example of such an integrated planning approach for this. This is described in [Sch18] and other papers. At IIT Bombay also, many parts of this activity have been developed over the years. Some of these resources are indicated at [Bel].

A landmark study to do with timetabling and railway operations as a whole is the complete revamp of the Netherlands Railway timetable (including sectional timings, terminals, rolling stock and crew) [Kro09]. This significant plan was brought about by many years of cooperative work between railway personnel, some of whom had university appointments, academics and software professionals. The core of an ecosystem and skills for such a concerted effort for the admittedly much larger railway system in India exists and should be cultivated, as the impact can be significant.

# References

[Alf06] A. Alfieri, R. Groot, L. Kroon and A. Schrijver. Efficient circulation of railway rolling stock, *Transportation Science*, 40(3):378–391,2006

[Bel] https://www.ee.iitb.ac.in/~belur/railways/

[Cap07] A. Caprara, L. Kroon, M. Monaci, M. Peeters, and P. Toth. Passenger railway optimization, *Handbooks in Operations Research and Management Science*, 14:129–187, 2007

[Jai19] D. Jain, Mumbai Suburban Railway Timetabling, *M.Tech Dissertation,* Department of Electrical Engineering, IIT Bombay, 2019

[Kas19] N. Kasliwal, P. Sudarshan, M.N. Belur, N, Rangaraj, Crew Planning for Commuter Rail Operations, a case study on Mumbai, India, *Technical Report,* 2019

[Kro09] L. Kroon, D. Huisman, E. Abbink, P.J. Fioole, M. Fischetti, G. Maroti, A. Schrijver, A. Steenbeek, R. Ybema, The new Dutch timetable: the OR revolution, *Interfaces*, vol. 39, no.1, pages 6-17, 2009

[Lan05] A. Landex, and A.H. Kaas, Planning the most suitable travel speed for high frequency railway lines, *Proceedings of the 1st International Seminar on Railway Operations Modelling and Analysis*, Eds. I.A. Hansen, F.M. Dekking, R.M.P. Goverde, B. Hindergott, L.E. Meester, The Netherlands, 2005

[Lin] https://www.lintim.net

[Low04] S. Low, F. Paganini, and J. Doyle, "Internet congestion control," *IEEE Control Systems Magazine*, vol. 22, no. 1, pages 28-43, 2004

[Mar04] D. Marx. Graph colouring problems and their applications in scheduling. *Periodica Polytechnica Electrical Engineering*, 48(1-2):11-16, 2004

[Mig17-1] D. Miglani, *Simulation and Analysis of Mixed Traffic in Railway Junctions,* M.Tech. Dissertation, Department of Mechanical Engineering, IIT Bombay, 2017

[Mig17-2] D. Miglani, M.N. Belur and N. Rangaraj, Railway junction simulation and analysis for mixed rail traffic, *Proceedings of the Symposium on Advanced Train Control and Safety Systems for Indian Railways (ATCSSIR)*, Kharagpur, May 26-27, 2017

[Odo87] A.R. Odoni, The flow management problem in air traffic control, *Flow Control of Congested Networks,* Springer, pages 269-288, 1987

[Ope17] Operations Document, Indian Railways, National Academy of Indian Railways, (Last accessed: Oct 2017). www.nair.indianrailways.gov.in/uploads/files/1430369423822-operating.pdf

[Rag86] G. Raghuram. Study of the problem of improving the utilization of the non-suburban passenger fleet in indian railways. *Working Paper No.600*, Indian Institute of Management Ahmedabad, 1986

[Ran02] N. Rangaraj and B. N. Vishnu, Node capacity and terminal management on Indian Railways, In: *Vision 2025,* Vadodara Railway Staff College, India, 2002

[Ran06] N. Rangaraj, M. Sohoni, P. Puniya, and J. Garg. Rake linking for suburban train services, *Opsearch*, 43(2):103–116, 2006

[Ran18] N. Rangaraj and M.N. Belur, A concept note for railway timetabling to rationalize and improve capacity utilization, *NITI-Aayog report*, January 2018

[Sal17] S. Salsingikar, N. Rangaraj and L. Sathishkumar, Analysis and planning of train movements at a railway junction, *Proceedings of the International Conference on Railway Transport Technology,* Lille, France, April 4-7, 2017

[Sch18] A. Schiewe, S. Albert, J. Pätzold, P. Schiewe, A. Schöbel, J. Schulz LinTim: An integrated environment for mathematical public transport optimization. Documentation.Preprint-Reihe, Institut für Numerische und Angewandte Mathematik, Georg-August Universität Göttingen, 2018

[Sub19] S. Subramanian, N. Rangaraj, The effect of rolling stock standardization on rake utilization and punctuality of services, *Technical Report,* IEOR, IIT Bombay, 2019

[Uic04] UIC leaflet 406, Capacity, *UIC International Union of Railways*, France, 2004

[Vid17] R. Vidyadhar, S. Dutta, N. Rangaraj and M.N. Belur, Simulation and analysis of mixed traffic on railway sections, *Proceedings of the Symposium on Advanced Train Control and Safety Systems for Indian Railways (ATCSSIR)*, Kharagpur, May 26-27, 2017