School of Computing
FACULTY OF ENGINEERING

UNIVERSITY OF LEEDS

Intermediate Report

Assessment Scoring For Train Timetable Design

Daniel Partington

Submitted in accordance with the requirements for the degree of
BSc Computer Science with Mathematics

2018/2019

Type of Project: ________Empirical Investigation_____________

The candidate confirms that the work submitted is their own and the appropriate credit has been given where reference has been made to the work of others.

I understand that failure to attribute material which is obtained from another source may be considered as plagiarism.

D.Partington

Table of Contents

List of References

[1]Performance indicators for railway timetables; Rob M.P. Ingo A.; Published in: 2013 IEEE International Conference on Intelligent Rail Transportation Proceedings

[2]An assessment of railway capacity;M.AbrilaF.BarberaL.IngolottiaM.A.SalidoaP. ; Published in April 2007

[3] Railway Track Capacity: Measuring and Managing; KhademSameni; Published in 2012

[4] Online train delay recognition and running time prediction; Ingo A. Hansen, Rob M.P. Goverde, Dirk J. Van der Meer; Published 2014

[5] Evaluation of railway capacity, Landex, Alex; Kaas, Anders H.; Schittenhelm, Bernd Hermann; Schneider-Tilli, Ja, Published 2006

[6] Estimation of Railroad Capacity Using Parametric Methods; SubhroMitra, Denver Tolliver, Sushil Mitra, Khalid Bachkar and PoyrazKayabas; Published 2010

[7]Transportation Research Part E: Logistics and Transportation Review; M.AbrilaF.BarberaL.IngolottiaM.A.SalidoaP.TormosbA.Lovab; Published September 2008

[8] N. Policella, 'Scheduling with uncertainty: A proactive approach using partial order schedules', PhD Thesis, UniversitadegliStudi di Roma: La Sapienza, (2005).

[9] Introduction to max-plus algebra; Hiroyuki Goto; Published 2014

1 Introduction

1.1 Introduction to the timetabling problem

The Great British Rail Network since its creation has continuously aspired to take strides in improving all aspects of the service it provides. One crucial feature of this service of course being the creation and delivery of timetables for train lines, and so naturally their aim is to ensure such timetables are of the highest standard possible, which poses the question: What makes a good railway timetable?

A railway timetable must be able to withstand delays, perturbations, and variations in operating conditions without losing functionality, so that the scheduled train services are actually achieved according to plan as best as possible[1]; All whilst ensuring the amount of passengers / freight able to be transported over each given route is corresponding to the current demands of society and business'.

As it stands, despite guidance and heuristics from experienced planners in designing 'good' timetables, there are no recognised measures or lead indicators of whether a timetable will be performant.[Need referencing since it was from the tracsis work spec?]

In the following report I will propose characteristics that affect the quality of a timetable and explore ways to evaluate them using only data which is available prior to the execution of the timetable.

1.2 Motivation

Being able to evaluate a timetable before execution and give with a degree of assurance some kind of indication as to how successful that timetable is likely to be would be an invaluable ability in the process of timetable design, enabling one to consider a prototype and identify areas of weakness or strength with the capacity to act on these assessments. With the aspiration to achieve such a feat it is crucial to first consider the factors which contribute to a 'good' timetable.If it is possible to segregate such factors into independent categories are we then able to quantify how well a timetable stands in each segment in order to pull together some overall classification as to just how good it is? Over the duration of this project I will look to investigate what factors contribute to a successful timetable and explore methods in quantifying if or not a timetable possesses these attributes.

1.3Aim

The main aim of this project is to investigate the factors (if any) contributing to the quality of a timetable, and exploring if such factors exist, can they be quantified by some sort of mathematical method which will allow us to give a degree of indication of the calibre of a given timetable before execution.

Deliverables

- A Report detailing the research into factors contributing to timetable performance and the decision process of methodologies to quantifying them
- Research, deduce, and create various mathematical methods of quantifying my chosen factors and an overall scoring system considering all of the factors.

Background Research

3.1 Railway Capacity Utilisation

3.11 Introduction to Railway Capacity

The Capacity of a railway is a seemingly easy but rather inaccessible concept that has been a hot topic of discussion in terms of its definition within the rail industry for many years. Although capacity seems to be a self-explanatory term in common language, its scientific use may lead to substantial difficulties when it is associated to objective and quantifiable measures. It is a complex term that has

"Capacity is the number of trains in a given time period that a rail line can accept without exceeding a specified limit of queuing time." (Peat Marwick and Partners, 1977)[3] numerous meanings and for which numerous definitions have been given. When referring to a rail context, it can be described as follows:

Definition: Railway Capacity

The idea of being able to quantify this concept of a capacity for railways gives rise to a number of different 'types' of capacities as such, which are listed and described throughout the remainder of this section.:

Definition: Theoretical Capacity

Maximum (Or Theoretical) capacity is the number of trains that could run over a route, during a specific time interval, in a strictly perfect, mathematically generated environment, with the trains running permanently and ideally at minimum headway (i.e. temporal interval between two consecutive trains). [2]

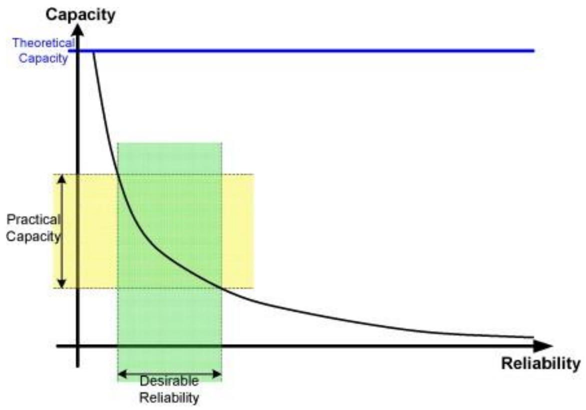
The Theoretical Capacity therefore serves as an upper limit for line capacity. It assumes that traffic is homogeneous, that all trains are identical, and are evenly spaced throughout the period with no disruptions. It ignores the effects of variations in traffic and operations that occur in reality, meaning a maximum throughput can be obtained for an optimal configuration of routes and speeds of trains with in particular homogeneous traffic where all successive trains have exactly the same optimal speed and stop pattern. In reality however reaching such a capacity is unattainable, due to the natural variances in the optimal speeds, orders and frequencies of trains within a timetable (Timetable pattern).

Taking into account these factors enables us to define a new type of capacity, the Fundamental Capacity, which can be thought of as the practical limit of "representative" traffic volume that can be moved on a line at a reasonable level of reliability. The "representative" traffic reflects the actual train mix, priorities, traffic bunching, etc.

Definition: Fundamental Capacity

The fundamental(Or practical) Capacity is the maximum number of trains per time period that can be operated given the timetable pattern (train types, frequencies, orders, and speeds). [2]

If the theoretical capacity represents the upper theoretical bound, the fundamental capacity represents a more realistic measure. Thus,thefundamental capacity is calculated under more realistic assumptions, which are related to the level of expected operating quality and system reliability, as shown in Fig. 1. It is the capacity that can permanently be provided under normal operating conditions. It is usually around 60–75% of the theoretical capacity, which has already been concluded by Kraft (1982). Practical Capacity is the most significant measure of track capacity since it relates the ability of a specific combination of infrastructure, traffic, and operations to move the most volume within an expected service level.[3]



Figure(1)Practical capacity involves the desirable reliability [2]

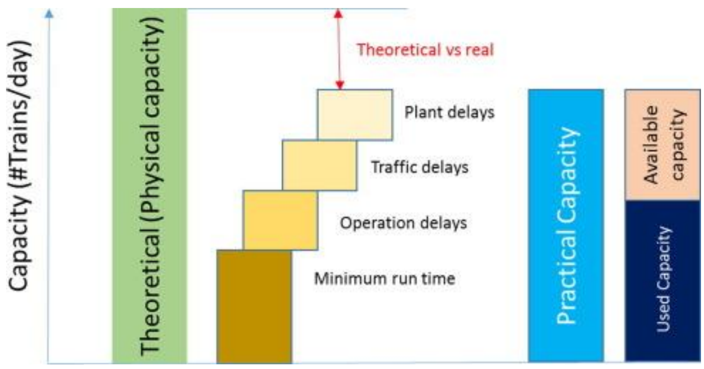
Definition: Used Capacity
Used capacity(Or Infrastructure Occupation) is the actual traffic volume occurring over the network. It reflects actual traffic and operations that occur on the line. It is usually lower than the practical capacity. [2]

Given an infrastructure and a range of train types, the amount of used capacity is determined by the specifications of the timetables over the tracks. The ability of the timetable planners to have this certain degree of control over the used capacity value opens up the challenge of finding the optimal proportion of the Practical Capacity to use in order to create timetables of maximised efficiency. The recent history of this problem including the factors affecting it and the approaches taken to tackling it are found in the subsequent section.

Definition: Available Capacity
Available capacity: It is the difference between the Used Capacity and the Practical Capacity. It is an indication of the additional traffic volume that could be handled in the route. If it allows new trains to be added, it is a useful capacity; otherwise, it is 'lost' capacity. [2]

Available Capacity = Theoretical Capacity – Used Capacity

The general hierarchy of these types of capacity and their distribution is depicted in Figure 2 Below



Figure(2) An Example shared Capacity [3]

3.11 The Railway Capacity Challenge

The main concept of transportation is moving passengers and goods from one place to another. Transportation affects different aspects of human lives from daily individual level to long-term socio-economic welfare and sustainability of societies, and the transportation of such passengers and goods should be performed in an efficient, safe, secure and reliable manner with the lowest external and internal costs possible.

However, road congestion, higher fuel costs, rising incomes, privatisation of railways along with concern for sustainability of transport have resulted in enormous growth in rail passenger and freight in the last decade. Growth in railway demand is one side of the coin; the other side of the coin is limited railway infrastructure. Many railways around the world now find themselves facing a challenge to accommodate necessary train services on their infrastructure, due to the capacity of infrastructure not increasing proportionately to keep up with the pace of demand of increased passengers and freight. Therefore, a so-called “railway capacity challenge” has emerged which is schematically depicted in Figure(3).

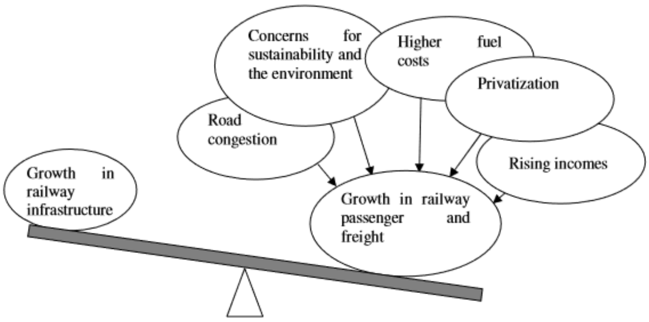


Figure (3) Analogy of factors influencing the challenge

To give evidence of this, statistics in Europe show a total increase of 32% in tonne/km of freight transported by rail and an 9% increase in rail passenger-km between 2001 and 2010 (UIC, 2011c) . During this time, railway infrastructure occupation has increased by just 5%.[3]

Note: The method of how to quantify this infrastructure occupationis covered in the next section

Year	**2001**	**2010**	**2001**	**2010**	**2001**	**2010**
Item	Passenger-km (billion)		Tonne-km (billion)		Line-km	
Value	575.3	626.2	1861.0	2,454.4	353,170	370,387.9
Growth (2010/2001)	+8.86%		+31.88%		+4.88%	

Figure(5) Growth in rail passenger, freight and infrastructure across Europe. Data source:(UIC, 2011c)

The emergence of the railway capacity challenge in recent years has made the objective of utilising the capacity of rail sections an increasingly difficult one. Growing demands of freight and passenger transportation without a simultaneous and corresponding increase to the size of railway capacities means the only way to accommodate for the increase in demand for a given infrastructure would be through scheduling more rolling stock, which in turn will increase the Used Capacity to a value closer to the practical upper bound of the Fundamental capacity. However just because there is capacity available to be used doesn't necessarily mean it will make the transportation within the railway timetable more efficient. In fact, using too high of a proportion of the fundamental capacity means very heavy, compressed train flows leaving limited resources to recover from delay incidents; meaning small delays can go on to have detrimental knock-on effects right throughout the timetable. The flip side to this would be to utilise a very low portion of the Fundamental capacity, meaning a low volume of train-flow, however with a good amount of resource to react and recover to incidents causing delays; an arguably more stable approach, however it doesn't tackle the issue of the railway capacity challenge. The table in Figure(6) below displays these general ideas about the potential consequences of using a range of different infrastructure occupations

	Level of service (LOS)	**Description**		**Volume/capacity ratio**
	A	Below Practical Capacity	Low to moderate train flows with capacity to accommodate maintenance and recover from incidents	0.0 to 0.2
	B			0.2 to 0.4
	C			0.4 to 0.7
	D	Near Practical Capacity	Heavy train flow with moderate capacity to accommodate maintenance and recover from incidents	0.7 to 0.8
	E	At Practical Capacity	Very heavy train flow with very limited capacity to accommodate maintenance and recover from incidents	0.8 to 1.0
	F	Above Practical Capacity	Unstable flows; service breakdown conditions	> 1.00

Figure(7)Level of service grades (LOS) (Cambridge Systematics, 2007)

This again poses the question: For a given infrastructure, what is the optimum amount of capacity to utilise?

In order to tackle railway capacity challenge, efficient utilisation of railway infrastructure is critical as building new railway lines is extremely costly and time-consuming. Compared to road transportation which also has limited practical capacity on its main infrastructure, the concept of railway capacity is not well explored, and so in the following sections I will be looking into methods in measuring Capacity Utilisation and subsequently investigating the effects of the different utilisations on the efficiency of the execution of a timetable.

3.12 Quantifying Capacity Utilisation

A number of different approaches to evaluating railway capacity and its utilisation have been developed over the past 2 decades, and with the concept of capacity being applied to railway infrastructures being a relatively new one, there isn't a general consensus throughout the railway business as to which method is best. Each method takes a different perspective in tackling the problem and can be categorised as follows:

- Analytical methods
- Parametric models
- Optimisation technique
- Simulation

When eventually coming to investigate approaches to scoring timetables in a way that gives some degree of indication of its quality and likelihood to perform I will only take forward the Analytical approach, and shall cover the reasons for this decision at the end of the section after briefly reviewing each of the above.

3.121 Analytical methods – The UIC 406 Method

Analytical methods take mathematically based theories in order to deduce equations that quantify the maximum number of trains for a line section without timetable. There are a number of analytical methods put into practise, however the most famous and commonly used is the UIC 406 method.

The UIC 406 method provides a straightforward method of timetable analysis and used capacity by compressing the timetable so that the buffer time is zero (UIC, 2004). As recommended by this standard, ideally each line section used for compression should be reduced to the line section between two neighbouring stations (without overtaking or crossing possibilities). Firstly the timetable is produced. Then the railway network is divided into sections at:

- Junctions
- Change of train order
- Change in number of trains
- Change in number of tracks
- Overtaking and crossings station

The eventual capacity calculation is based on the compression of timetable graphs on a defined line or line section, and so the next step is to perform this compression. All single train paths are pushed together to the minimum headway time, so that no buffer times are left. The compression of the timetable graph has to be done with respect to the train order and the running times. This means that neither the running times, running time supplement, dwell times or block occupation times are allowed to be changed. Furthermore, only scheduled overtaking and scheduled crossings are allowed. An example of timetable compression can be seen in Figure (8) [5]

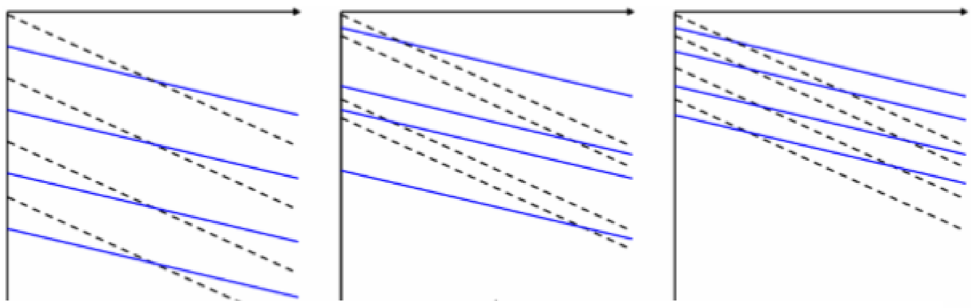
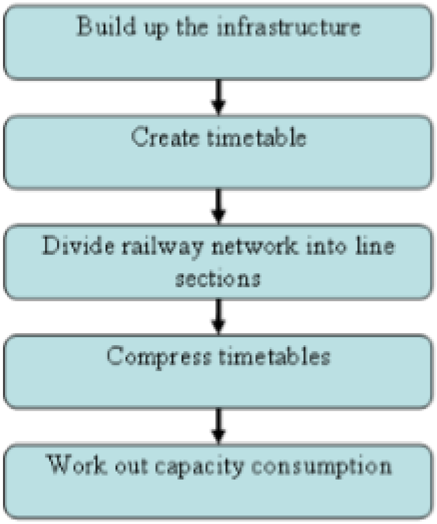


Figure (8): Actual timetable for quadruple track (a) and the same timetable compressed keeping the train order (b) and optimizing the trainorder (c)

Once the timetable has been compressed sufficiently, the next, and final phase is to carry out the appropriate calculations to obtain capacity consumption.

The general workflow of the UIC method is illustrated in the workflow displayed in Figure(9).



Figure(9) A simple flow chart describing the UIC406 Method

The UIC formula for determining used capacity is then:

k=A+B+C+D (1)

k: Total used time (min)
A: Infrastructure occupation (min)
B: Buffer time (min)- the time added to decrease the risks of delay propagation
C: Supplement for single-tracks (or crossing time)(min) – the time added to reduce size of delay
D: Supplements for maintenance (min)

K= k×100 / U (2)

K : Used capacity (%)
U: Chosen time window (min)
(UIC, 2004)

Out of the 4 input parameters in equation (1), all are easily extracted from the compressed version of the graphical timetable, apart from B, the buffer time. Whichaccording to capacity utilisation ratio and headway, can be found as follows:

$$K_{\max} = \frac{\Delta T}{t_{h\min}}$$

$$K_f = u \times k_{\max}$$

$$K_f = u \times \frac{\Delta T}{t_h}$$

K_f : Usable capacity (theoretical capacity) (number of trains)
$K_{\max}$: Maximum capacity (practical capacity) (number of trains)
ΔT : Observation period (min)
u : Percentage of utilisation of the maximum capacity (used capacity)
$t_{h\min}$: Minimum headway (min)
t_b: Buffer time (min)

$$K_f = \frac{\Delta T}{t_{h\min} + t_b}$$

By inserting the second formula into the first one, the buffer time (min/train) would be:

$$K_f = u \times \frac{\Delta T}{t_h} = \frac{\Delta T}{t_{h\min} + t_b} \Rightarrow t_b = \frac{\Delta T}{K_f} - t_{h\min}$$

3.122 Parametric methods

Parametric models use a selection of parameters of railway infrastructure and operation to describe and analyse capacity utilisation. The Parametric approach to railway capacity is a relatively unexplored one, with the first ever model only being made in 1975 by Prokopy and Rubin. This model laid down the foundations for all subsequent parametric models developed, with all of them using the same general approach with the following five steps[6] :

- Modification of Prokopy and Rubin's (1975) train dispatching simulation (TDS)
- Identify key parameters affecting capacity
- Procedure for parametric analysis
- Evaluation of the parameters
- Validation of the model and verification of the accuracy

The main benefit of the parametric approach is that one is able to give a sophisticated assessment of capacity without getting engaged in the actual simulation; And although the parametric method has been implemented successfully on many occasions, perhaps most noticeably by Krueger (1999) for the Canadian National Railway, I decided against taking it any further in my investigations due to two main reasons:

1. The assessmentdoesn't result in a definitive numerical figure, meaning it would be harder to take any outcomes further when combining with the results of the other sub-catergories
2. The Parametric approach is more appropriate for the railways where the operation manager and the infrastructure owner are the same entity in order to include these different parameters in one model

3.123 Optimisation Technique

Due to the very complex and multidisciplinary nature of railway capacity, mathematical programming and operations research are not directly used for modelling and optimising capacity utilisation (i.e. maximising efficient capacity utilisation, subject to demand/infrastructure/ signalling/ operational/ rolling stock/ fare/ access charge constraints). Therefore, optimisation techniques are extensively used only for subproblems of capacity utilisation especially train scheduling, rescheduling and routing as well as track and platform allocation. It is due to this reason along with the fact that optimisation methods need to be adapted to each application environment(which isn't ideal since I am seeking to find more generalised results), that I decided against taking this approach any further.Recall that capacity utilisation is just one sub category of the performance indicators I will be exploring, and so I deemed this approach unnescarily complicated for the purposes of this project. There have however been interesting developments in this field over the past 30 years which could prove significant in a study perhaps more solely focused on capacity. A brief overview of these findings are listed below:

Assad (1980) surveyed different mathematical models for optimising railway operations. Cordeau et al. (1998) review train scheduling and rescheduling in their comprehensive survey. Tornquist (2005) provides an overview of research in the field of railway scheduling and dispatching. Hansen et al. (2008) present chapters about state-of-the-art techniques on timetable design principles, infrastructure modelling, timetable stability analysis, optimisation models for railway timetabling, simulation, rescheduling and performance evaluation. Lusby et al. (2009) provide a recent survey on track allocation models and methods.[3]

3.124 Simulation Methods

A simulation is the imitation of an operation of a real-world process or system over time. It is the representation of dynamic behaviour of a system by moving it from state to state in accordance with well-defined rules. Simulation methods provide a model, which is as close as possible to reality, to validate a given timetable. For train scheduling, simulation has often been used in combination with other methods, originating what could be defined as "hybrid models". Since the 1980s, Petersen (1974) are quoted for their work, which uses combined techniques, such as dynamic programming and branch-and-bound in a simulation context. A composite simulation and optimization method also appears in the work of Jovanovic and Harker (1991).[7]

Besides purely academic products, various simulation environments have been produced and are commercially available and utilised within the rail industry. The general performances of these simulators are comparable, with their main technical differences concerning interface design, user interaction and flexibility, track infrastructure, other data management processes, and integration with company information systems. These tools normally generate timetables by time-stepping simulation using train motion differential equations. Alternatively, these tools can be used to validate a given timetable, which is provided by optimization methods. They can detect delays and analyze interferences on the preliminary timetable.

Although simulations provide a unique visual perspective on capacity utilisation, the analysis of the corresponding outputs will be somewhat subjective to an extent. Simulation methods can be very efficient in highlighting areas in the timetable perhaps more susceptible to interferences or delays for example, however yet again there are no definitive quantifications of capacity utilisation directly obtainable from simulations. For this reason I ruled out using such methods in the experimentation stage of my project.

Summary of methods and choice of UIC Method

3.2 Timetable Feasibility

3.2.1 Introduction to timetable feasibility

The concept of feasibility with regards to timetabling is the assessment of whether a given timetable is actually achievable when executed in reality, this assessment is made by evaluating two key factors:

- the calculated minimum timeequalling or exceeding the time designated by a timetable for a train to run from one station to another.
- If all trains are functioning as intended for a given timetable, then no two trains should interfere with one another. [1]

With this said, before giving a concrete definition of the notion, it is necessary to first introduce the following terms:

- **Process time:** Given two locations, A and B, the Process time of a train transporting from A to B is the total amount of time taken from the point at which the Train first arrives at A, to the point it first comes to a standstill at B – so it is important to note that the process time is not just simply the time taken for a train to travel from point A to B, but also takes into consideration the time taken in other states such as dwell time, turn-around time, etc.

- **Conflict free:** Two train paths are said to be 'Conflict free' if and only if none of the trains on such paths will ever have to brake due to a conflicting train route, so long as all trains adhere to their specified schedule

With the two terms specified and explained above, we are now ready to define the concept of feasibility with regards to train timetables:

Definition: Timetable Feasibility

Timetable feasibility is the ability of all trains to adhere to their scheduled train paths. A timetable is feasible if the individual processes are realizable within their scheduled process times and the scheduled train paths are conflict free [1]

3.2.2Instances of Unfeasible Timetables

One may think that it is a given that all timetables should possess this characteristic of feasibility and that only poor timetabling design could result in a timetable becoming unrealisable. However, with the increasing need for timetables of a given infrastructure to be as efficient as possible in transporting passengers and goods, even experienced timetable planners have been known to produce timetables which prevailed to be unfeasible, resulting in delays when it comes to executing the schedules. This willingness to squeeze as many journeys as possible into the infrastructure to the point at which even perfect execution of the timetable leads to delays is evidence of the drastic measures Train Service Companies are prepared to take in order to tackle the railway capacity challenge detailed in section 3.1.2. Proving once again the importance and usefulness of being able to give sophisticated evaluations of timetables before execution;So that such occurrences can be identified and rectified without significant effects on the overall efficiency of the timetable.

There are a number of reasons why a timetable may become unrealisable, with the two most common of being:

- **Rounding Errors**: A timetable may become unrealisable and thus infeasible if scheduled departure times (at short stops) are rounded down to whole minutes. This rounding prevents that trains have to wait for their scheduled departure time since early departures are prohibited. Such rounded departure times must be compensated for downstream to guarantee feasibility of an aggregated train run between two main stations. This is essential for stability. In general, a commercial timetable with a precision of a minute as published to travellers may not be (locally) realizable with respect to the rounded event times. However, such a commercial timetable may have an underlying working timetable with a higher precision that is feasible.

- **Human Mathematical Errors:**The calculation of minimum process times isn’t a trivially simple one, in fact it can become rather complex when considering each fundamental stage contributing to the overall process time.Like any other calculation it is naturally susceptible to errors such as mistakenly excluding parameters or the miscalculation of contributing values relating to particular calculations.

Feasibility is therefore a necessary condition to check when evaluating the quality and performance of a timetable. Its evaluation can be thought of as the first the first point of action when assessing a timetable, in the sense that we can assure ourselves with a strong degree of certainty that once a timetable is deemed‘feasible’,as long as everything goes to plan there shall be no conflicts experienced. In reality, however, this is obviously not always the case - delays occurr regularly for various reasons, meaning the ability of a timetable to react to said delays is another crucial assessment to make when predicting how successful a timetable will be(Covered in later sections). For now however, I will focus on methods ofassessing the initial feasibility of train timetables.

3.2.2 Quantifying Feasibility

Unlike Capacity Utilisation with the UIC 406 Method in the previous section, there isn’t a generally preferredmethod across the industry which is conformed to whenassessing feasibility. There has however been some basic approaches to evaluating feasibility proposed, such as counting the number of train path conflicts for example. Whilst these methods aren’t necessarily considered a wrong way to estimate feasibility, there exists no records of them being implemented in the development phase of producing railway timetables – perhaps due to their very basic and intuitive nature.

In order to meet the objectives of this project, I require a way of evaluating feasibility that results in a numerical representation of the characteristic. Subsequently, sinceit became apparent that implementing pre existing methods already proven to give both successful and accurate results wouldn't be possible, I formulated my own. In devising this method I wanted to ensure that both factors constituting the definition of feasibility were incorporated, and so before introducing my approach, I first recall that a Timetable may be deemed feasible if the following two conditions are met:

1. **Scheduled times Exceed Process times**
2. **Train Paths are conflict free**

But instead of simply just categorising timetables as feasible/unfeasible based on whether they meet the above conditions, I wanted to use information on both the actual trains running and the timetables themselves to not only determine whether a timetable is feasible or not, but give an numerical indication of how comfortably feasible they are. For example, if all train journeys in timetable A are only scheduled 1 minute more than each of their respected process times, but timetable B allocates 10 additional minutes for each process time to the scheduled times, then although both timetables meet condition 1 above, timetable A is only just realisable where as timetable B is comfortably so.
With the intentions of representing this feature, I propose the following as a method to quantify each factor contributing to feasibility:

EITHER CONDITION FAILS => INFEASIBLE

BOTH CONDITIONS MET => APPLY FOLLOWING CALCULATION:

CONDITION 1:
AVERAGE PERCENTAGE DIFFERENCE BETWEEN SCHEDULED AND PROCESS TIMES FOR EVERY JOUNREY IN THE TIMETABLE

CONDITION2:
AVERAGE TIME OF HOW CLOSE ANY TWO CONSECUTIVE TRAINS SHARING TRACKS ARE TO CONFLICTING

This will result in 2 numerical figures for each timetable assessed, where in both cases the higher the figure, the more comfortably feasible the timetable is. These numerical figures can then be taken forward and compared with the results obtained from the other sub categories considered, as initially intended for this project.

3.2.3 Pre-assessment of method

Of all the performance indicators considered in this project, the concept of a timetable possessing feasibility is perhaps the most basic, and certainly the most intuitive. For this reason I think utilisinga method of quantification matching the concept in complexity is an appropriate approach to tackling the problem. A method too complex I believe would only unnecessarily overcomplicate the problemand may even lead to less accurate results.

With this said I believe that although my proposed method is indeed basic and easy to carry out as desired, this doesn't necessarily imply that it is without fault. For example, my chosen approach will assume that both contributing factors influence feasibility equally. However in reality it may be that scheduled times being closer to the threshold of infeasibility isn't as much of an issue as perhaps routes almost conflicting, or vice versa. Also, it is important to remember that this is a completely untested method – after experimenting with a number of alternate techniques on various timetables it may become apparent that there is in fact other, more efficient ways to quantify feasibility. However, due feasibility being only one of 5 performance indicators to be explored, I deemed this experimental stage with alternate methods outside of the scope.

3.3 Timetable Stability

Introduction to timetable stability

Delays to train timetables to some extent are somewhat of an inevitability in reality. No matter how good an initial timetable design is, at some point over time it is highly likely that there will be an external factor outside of reasonable control of the train operators that causes delays to the planned schedules. Even the very accomplished and efficient Japanese railway system once famously in November 2017 felt it was necessary to make a public announcement for a train in Tokyo arriving just 20 seconds late! Saying they "sincerely apologise for the inconvenience caused", leaving the rest of the world envious of such high standards. But the point is that delays to train timetables are a part of life, and once this inevitability is accepted, it is crucial for timetable design to take steps in incorporating features to ensure to the best of their ability that when such delays occur, they have the resources to react and rectify the issue as quickly and as efficiently as possible. Possessing this ability of coping with delays that we refer to as stability.

As in previous sections, before introducing a definition to the concept I will run over a couple of useful terms that I shall use throughout this section:

- **Primary/InitialDelay:**The primary delay is the first delay caused in a given line section - caused by variations in a process by which the process takes longer than its scheduled process time
  Note: Initial delays include departure delays at the origin station and delays of trains entering a network from the outside (such as at country borders).

- **Secondary Delay:** Secondary delays to a train are delays caused by the propagation of one delayed train to another.

The explanation of the above terms above enables us to now give a more concrete definition to stability:

Definition: Timetable Stability

Timetable stability is the ability of a timetable to absorb initial and primary delays so that delayed trains return to their scheduled train paths.[1]

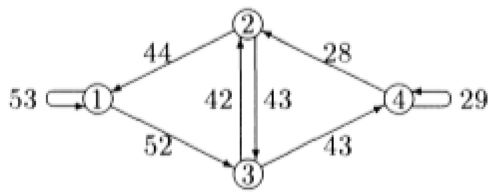
Hence, with this definition a timetable is ‘stable’ if any train delay can be absorbed by the time allowances in the timetable without active dispatching. A timetable in reality can therefore never really be thought of as completely stable, since there is no cap on the size of the delay within the definition. I will therefore seek ways in which we can give some sort of rating to the stability of a timetable in order to give a degree of indication as to how well a given timetable will be able to react to primary delays in order to minimise secondary delays, which I shall cover in the following sections..

Method proposal

Periodic railway timetables constitute of train lines operating at regular intervals and synchronization at transfer stations that enable transfers between train lines. This periodicity can be exploited in a discrete event dynamic system(DEDS) by modelling event times as a function of the period, enabling efficient analysis to be carried out on the stability of the timetable by employing Max plus algebra, a discrete algebraic system which allows a periodic timetable to be represented a linear system of equations. The steps to carrying out this method are as follows:

1. Represent the timetable as The Precedence Graph

The precedence graph is a directed graph consisting of a set of n nodes and a set of m arcs. The nodes corresponds to train departures at transfer stations, and the arcs represent routes between the stations. A weight a_{ij} is assigned to each arc (i, j) corresponding to the sum of the running time t of train i and the connection time (either stopping time or transfer time depending on whether its a feeder train or not) from train i to j.
e.g



Note, We assume that the graph is strongly connected, i.e., there is a (directed) path between any node i to any node j.
From the graph we can define the following:

1. **Circuit** - a (closed) path where the end points coincide.
2. **weight of a circuit** - the sum of its arc weights
3. **circuit length** - the number of arcs (trains) in the circuit

4. **The cycle mean** – weight of circuit / circuit length
5. **critical circuit** - a circuit with maximum cycle mean
6. **Critical cycle time** - The maximum cycle mean

2. Interpret the graph as a DEDS

Once the prescedence graph of the timetable concerned has been constructed, we can use the details to compose constraints on the departure times based upon earlier journeys within the system:
Let $x_i(k)$ be the the kth departure time of train i. We are concerned with a cyclic timetable. Therefore the departure times are periodic recurrent events. Let k be a counter denoting the current period. Let $d_i(k)$ be the scheduled departure time of train i in period k.

The time at which train i departs is then:

$$x_i(k+1) = \max\left(a_{i1} + x_1(k), \ldots, a_{in} + x_n(k), d_i(k+1)\right), \quad (1)$$

with the values of a_{ij} being equivalent to the weight of the directed arc between nodes j & i,, or if no such arc exists between the nodes then it takes the value of $-\infty$.

The above equation implies given an initial set of departure times for each x_i, the evolution of the rail system is completely dtermied, i.e., the subsequent departure times of all trains are uniquely fixed, describing a Discrete Event Dynamic Systems (DEDS) - The discrete events being the departures of trains at discrete times, and the dynamic behaviour coming from the fact the maximum element could in theory be any in the list of possible choices.

Equation (1) consists of two operations, addition and maximisation, and therefore in order to take the analysis any further we first must convert the equation by introducing new notation so that it may be manipulated as a system of linear equations.

3. Max Plus

Max-plus algebra is a discrete algebraic system in which the max and plus operations are defined as addition and multiplication in conventional algebra.[9] The set of elements considered in the max- plus algebra are the real numbers and the additional element $e = -\infty$.
To demonstrate this consider the example below:

$$a + b = \max(a,b)$$
$$a \times b = a + b$$

Thus, by switching to a Max Plus system obeying the two definitions above and ignoring the scheduled departure time constraint, equation (1) becomes:

$$x_i(k+1) = \bigoplus_{j=1}^{n} (a_{ij} \otimes x_j(k)), \qquad i = 1, \ldots, n, \qquad (2)$$

where $\oplus_{j=1}^{n} x_j = \max(x_1, \ldots, x_n)$ denotes repeated maximization. In vector notation eqn (2) is written as:

$$x(k+1) = A \otimes x(k),$$

where x(the state vector) = (x_1, . . . ,x_n) and A(the state Matrix) is the square (nxn) matrix whose ij$_{th}$ entry is a_{ij}.
This is now a linear system of equations in max plus algebra as required, with the addition of vectors defined component wise as it is in regular linear algebra.

4. <u>Stability Analysis by solving the eigenvalue problem.</u>

We first recall that The critical cycle time of railway system is the minimum cycle time for which a realizable timetable exists. Now consider the matrix A in equation (3), it can be proven that the critical cycle time and timetable of the railway system are equivalent to the eigenvalue and eigenvector of the A matrix of eqn (3), respectively.
Meaning that the values of λ and v which are the solution of:

$$A \otimes v = \lambda \otimes v.$$

Correspond to the eigenvalue, which is the critical cycle time for which a cyclic timetable exists and the eigenvector v, a timetable for which this critical regular behaviour is satisfied.

<u>Theorem</u>

Given a timetable with periodic time, T. The the scheduled max-plus linear system(7)is stable if and only if
λ <T.

If λ = T the system is critical and if λ > T the timetable is unstable.

If λ > T the circuit system will never be able to recover from even the smallest disturbances and delays may even propagate to lines outside of the circuit. λ < T implies buffer times exist in the service network which allow delays to settle, however the closer λ is to T the longer it will take for the system to returns to on schedule.

Timetable Robustness

There are a number of different processes in railway transportation to be considered during the formulation of a timetable. In order to make these considerations each process is first represented by a deterministic minimum process time calculated using a model with basic parameters derived from measurements, estimated by experts, or determined by default normative values. It is usually the case however that the different process times are acquired via different methods – for example detailed models may be used for the more complex process times such as running times, where as dwell and transfer times are commonlydetermined through simple rules. Various variables go into the estimation of these process times, on one hand we have the non-changing aspects such as distance between stations, model of train, track traction etc.. and on the other we have factors of a more stochastic nature such as passenger volumes, driver behaviour and weather conditions – all of which are dynamic and therefore in order to carry out any calculations that involve these factors, predictions must first be made. However, no matter which procedure was carried out for which process, a timetable must still be valid if the input parameters for each part of the timetable design are slightly different from their real values on any given day.

The element of uncertainty going into the calculation of minimum process times means another potential cause for an unsuccessful timetable – for example if on any given day the volume of passengers on a particular route was significantly more than anticipated then naturally passenger changeover times will increase, which in turn might cause a delay.The idea of robust schedules consists in solutions that can tolerate this degree of uncertainty during execution. In other words, they should be able to absorb dynamic variations in the problem due to both external reasons (exogenous events), and internal reasons[8] - giving rise to the following definition:

Definition: Timetable Robustness

Timetable robustness is the ability of a timetable to Withstand design errors, parameter variations, and changing operational conditions. .[1]

Such variations and imprecision's are unavoidable in reality, making robustness a very valuable attribute for timetables to possess – but what is it that makes a timetable 'robust'? The factors contributing to this trait are the subject of the following subsection.

Charachteristics of Robust Timetables

Intuitively, the tardiness of trains can have a domino like effect – If one service is late any subsequent services to be delivered from the affected unit are also likely to suffer. This is what's known as a secondary delay. The ability to absorb initial delays so that secondary delays are minimised is known as timetable stability(detailed in section 3.4) – however like with nearly all issues faced, the prevention of a problem before it's happened is better than fixing it once it's already happened. Creating robust timetables is therefore very desirable in industry and so in this section I will present a couple of key factors affecting the robustness of a railway timetable and how they can be measured

1. distribution of margins

Strongly related to robustness is the amount of margin inserted in the timetable. Margins can be added to the runtime and stopping time to prevent trains from arriving late despite small delays. Headway margin is used between any two consecutive trains in the timetable which serve to reduce the knock-on delay effects. A disadvantage of margins is, however, increased travel times and increased consumption of line capacity.
Not only the amount of runtime margin, but also its allocation is important.
Adding runtime margin concentrated early on a given line section line means that early appearing delays will not spread further down the line. However, if the disturbances occur

later on the line, the runtime margin added prior to the occurrence of the disturbance may become redundant. For this reason it is commonly seen in practise that margins are added into the middle of journeys along a line. Kroon et al., 2007[add ref], Fischetti et al., 2009[add ref], Vromans, 2005[add ref] give a measurement of the spread of margins, using the weighted average distance(VAD) to calculate the relative distance to the runtime margin from the start of the line in order to quantify the allocation. The method is carried out by dividing a line into N sections and letting st denote the amount of margin associated with section t, the 'WAD' can be calculated as:

$$WAD = \sum_{t=1}^{N} \frac{2t-1}{2N^2} \cdot s_t .$$

WAD is therefore a relative number between 0 and 1, where WAD = 0.5 means that the same amount of margin is placed in the first half of the considered line as in the second half, whereas WAD < 0.5 means that more margin is placed in the first half.

2. Heterogeneity of trains

The Heterogenity of a timetable describes the variation of trains used to carry out the services which it provides. A set of trains varying greatly in characteristics such as type, stopping patterns and most importantly average speed would be classed as heterogeneous. On the contraire, if the trains are very similar in these traits listed, they would be considered as homogenous - the opposite of heterogeneous, and it is Homogenity through which robustness is increased. This is because working with trains all of very similar attributes means that minimum headway times fall and in turn buffer times increase. There are numerous ways to decrease the heterogeneity of trains, a few of which are listed below:

- Slowing down long-distance trains
- speeding up short distance trains
- inserting overtakings (to lessen the consequences of faster trains)

A proposed measure of heterogeneity is given by the sum of shortest headway reciprocals (SSHR), which considers the smallest headway hi- between each train *i* and any consecutive train using the same track section and is given by:

$$SSHR = \sum_{i=1}^{n} \frac{1}{h_i}.$$

Although both these measurements are very informative and useful in the development phase of a train timetable, neither are enough in their own right to give a strong representation of robustness, and so in the next section I shall introduce more general approaches.

There has however been algorithms developed which incorporate some of these ideas with the purpose of increasing a given timetables robustness – with some successful results for example Sweden – further information on these can be found in appendix

<u>Methods To Quantify Robustness</u>

<u>Definition 1</u>

Definition 1:
We define t-robustness of timetable x (Robust x,t)) as the percentage of disruptions lower than t time units that the timetable is able to tolerate without any modifications in traffic operations (crossing, overtaken, etc).

This analytical method is related to the definition 1 presented above. This method gives us a time dependent indicative measure of how good the timetable behavior is when unexpected disruptions occur. Here, we consider different percentages of robustness for different sizes of disruptions. A disruption of 3 seconds in a train can be absorbed by practically all timetables. However disruptions of 50 minutes in a train could affect a significant proportion of the fleet. Finally, disruptions of 200 minutes in several trains could generate cancelations or large disturbances in the timetable.

Thus, a method is given in the following formula:

$$R(x,t) = \frac{100 * Nabsorbeddisruptions(x,t)}{T*S} \quad (2)$$

<u>(T*S = total number of possible disruptions)</u>

```
Function Nabsorbeddisruptions(x,t)
Input: timetable x, matrix of crossing, matrix of buffer times, time t
Output: Number of absorbed disruptions
forall train-i do
    forall station-j do
        k=j;
        while(not crossing[train-i, station-k])
           total=total+buffer[train-i, station-k];
           k++;
        remaining[train-i, station-j]=total - t;
        if remaining[train-i, station-j]>0;
           absorved++;
    end
end
 return absorved;
```

For instance, given a timetable x, if user selects t=120 seconds, by using definition 1, R(x,120)=40% means that 40% of disruptions can be absorbed without any modifications in traffic operations such as crossing, overtaken, etc. However if t=180, then the percentage of robustness decreases R(x,180)<R(x,120) due to the fact that the probability of absorbing higher disruptions is lower.

This is the method I wish take forward.

<u>Other notable methods to quantify robustness</u>

$$R(x)=\sum_{T=1}^{NT}\sum_{S=1}^{NS}Buff_{TS}*\%Flow_{ST}*TT_{S}*NSucT_{T}*(NS-S)/NS \quad (1)$$

where: R(x) is the robustness of timetable
 x;T is the train; NT in the number of trains;
 S is the station;
 NS is the number of stations;
 BuffTS is the buffer time of train T at station S;
FlowST is the percentage of passenger flow in train T ant station S;
NSucTT is the number of trains that may be disrupted by train T ;
 and finally TTS is the percentage of tightness of track between station S,S+1 .

Thus, given a timetable with k tracks, the tightest track has TTi=k/k=1 , the second tightest track has TTj=k−1/k , and so on. For instance, give a timetable with 21 stations, there exist 20 tracks. The tightest track is labeled to 1; the second tightest track is labeled to 0.95, the third is labeled to 0.9, and so on. Finally, the loosed track is labeled to 0.05.

This method however is only useful for comparing two timetables with eachother, as opposed to a stand alone numerical value for robustness. And so for this reason I decided against taking it any further

Timetable Resiliance

Timetable resilience is the flexibility of a timetable to prevent or reduce secondary delays using dispatching (re-timing, re-ordering, re-routing). Timetable resilience thus considers the interaction between the timetable and real-time traffic management, and is in particular related to the flexibility of a timetable to real-time traffic management in case of bigger delays, such as changing the order of two trains over shared infrastructure when the first train is too much delayed. Note that secondary delays are the target of timetable resilience. Recall that a secondary delay is a delay due to a path conflict between two trains: this can be either a route conflict in which case one of the two trains must reduce speed in response to the signaling and possibly wait for a red signal, or waiting on a delayed feeder train to enable passengers to transfer.