Contributions on microscopic approaches to solve the train timetabling problem and its integration to the performance of infrastructure maintenance activities

Diego Arenas

Université de Valenciennes et du Hainaut-Cambrésis

Thèse de doctorat

Pour obtenir le grade de Docteur de l'Université de

VALENCIENNES ET DU HAINAUT CAMBRESIS

Informatique

Présentée et soutenue par Luis Diego ARENAS PIMENTEL

Le 14/12/2016

École doctorale Sciences Pour l'Ingénieur (SPI)

Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH)

Évaluation des Systèmes de Transports Automatisés et de leur Sécurité (ESTAS) - IFSTTAR

Institut de Recherche Technologique Railenium

Contributions d'un modèle microscopique à la résolution du problème de construction d'une grille horaire et à la planification des activités de maintenance de l'infrastructure ferroviaire

JURY

* Président du jury

M. Feillet, Dominique	Professeur, École des Mines de Saint-Étienne - France	Rapporteur
M. Pesenti, Raffaele	Professeur, Università Ca' Foscari Venezia - Italie	Rapporteur
M. Chevrier, Rémy	Chef de Projets, SNCF I&R - France	Examinateur
Mme. Elloumi, Sourour	Professeur, ENSTA ParisTech - France	Examinatrice
M. Artiba, Abdelhakim	Président de l'Université de Valenciennes - France	Co-directeur
M. Hanafi, Saïd	Professeur, Université de Valenciennes - France	Co-directeur
M. Rodriguez, Joaquin	Directeur de recherche, IFSTTAR - France	Co-directeur
Mme. Pellegrini, Paola	Chargée de recherche, IFSTTAR - France	Encadrante

Université de Valenciennes et du Hainaut-Cambrésis

Doctoral Thesis

Towards the degree of Doctor in Philosophy at the University of

VALENCIENNES ET DU HAINAUT CAMBRESIS

Computer Science

Presented and defended by Luis Diego ARENAS PIMENTEL

December 14th, 2016

Doctoral School of Engineering Sciences (SPI)

Laboratory of Industrial and Human Automation control, Mechanical engineering and Computer Science (LAMIH)

Evaluation and Safety of Automated Transport Systems (ESTAS) - IFSTTAR

Technological Research Institute Railenium

Contributions on microscopic approaches to solve the train timetabling problem and its integration to the performance of infrastructure maintenance activities

JURY

* President of the jury

M. Feillet, Dominique	Professor, École des Mines de Saint-Étienne - France	Referee
M. Pesenti, Raffaele	Professor, Università Ca' Foscari Venezia - Italy	Referee
M. Chevrier, Rémy	Project Manager, SNCF I&R - France	Examiner
Mme. Elloumi, Sourour	Professor, ENSTA ParisTech - France	Examiner
M. Artiba, Abdelhakim	President of the Université de Valenciennes - France	Co-director
M. Hanafi, Saïd	Professor, Université de Valenciennes - France	Co-director
M. Rodriguez, Joaquin	Research director, IFSTTAR - France	Co-director
Mme. Pellegrini, Paola	Researcher, IFSTTAR - France	Supervisor

Résumé

La plupart des systèmes ferroviaires subissent une demande croissante de capacité. Pour y faire face, il faut construire de nouvelles infrastructures ou exploiter plus efficacement celles existantes, notamment en définissant des grilles horaires optimisées. Dans la littérature, la plupart des approches de construction des grilles sont basées sur des représentations macroscopiques de l'infrastructure, ce qui peut conduire à des solutions infaisables ou inefficaces. En revanche, les approches microscopiques reposent sur une modélisation réaliste du système ferroviaire, ce qui garantit la faisabilité et l'efficacité des résultats. Néanmoins, en raison de leur complexité, l'utilisation de ces approches est généralement limitée à une seule gare. Malgré l'optimisation de la grille horaire, les travaux de maintenance peuvent avoir un fort impact sur les circulations des trains. En présence de maintenances, il peut donc être nécessaire de redéfinir la grille horaire pour assurer une exploitation efficace de la capacité. Nous présentons deux contributions principales sous forme de deux approches microscopiques : une pour la conception de grilles horaires et l'autre pour leur redéfinition en cas de maintenance. La deuxième est la première approche microscopique qui apparaît dans la littérature pour aborder ce problème tout en considérant des aspects comme les limitations temporaires de vitesse. Nous démontrons la validité de nos approches et leur applicabilité dans des scénarios réels. De plus, nous montrons que les approches microscopiques peuvent être utilisées pour traiter des zones de l'infrastructure contenant plusieurs gares.

Mots Clés

Grilles horaires des trains, redéfinition des grilles horaires, représentations microscopiques, maintenance de l'infrastructure, limitations temporaires de vitesse, programmation linéaire en nombres entiers (PLNE).

Abstract

Most railway systems experience a growing demand of railway capacity. To face this demand, either new infrastructure must be built or a more efficient exploitation of the existing one must be attained. Timetables play a determinant role in the efficient capacity exploitation. Most timetabling approaches in the literature are based on macroscopic representations of the infrastructure. This may lead to inefficient and in some cases, impractical solutions. Instead, microscopic approaches are based on more realistic modelling of the elements of the railway system. This guarantees the feasibility of the timetables while promoting an efficient capacity exploitation. However, due to their complexity, the scope of microscopic approaches is typically restricted to main stations. Despite the optimization of timetables, the performance of infrastructure maintenance may severely impact the trains' circulations in the network. Therefore, the timetable may have to be rearranged to ensure an efficient capacity exploitation. We present two main contributions in this thesis: First, a microscopic approach for timetable design. Second, a microscopic approach for timetable rearrangement to cope with maintenance. This is the first microscopic approach in the literature to tackle this problem while also considering specific aspects as temporary speed limitations. After a thorough experimental analysis, we demonstrate the validity of our approaches and their practical applicability in real life scenarios. In particular, we show that microscopic approaches can be used to tackle large areas of the infrastructure, including several stations.

Keywords

Train timetabling, timetable rearrangement, microscopic representation, infrastructure maintenance, temporary speed limitation, mixed-integer linear programming (MILP).

Contents

Chapter 1

Introduction

1.1 Context & Motivation

In recent years, a steady growth of the railway demand has been observed in most European countries. This trend is expected to rise in the forthcoming years, which will lead to an exploitation close to the capacity limits of the railway networks. This is particularly true during peak hours in railway infrastructures of large cities, where even saturation circumstances are possible. To face this demand growth, the infrastructure managers must either increase the exploitation efficiency of the existing infrastructure, or undertake a series of construction activities to extend the infrastructure and increase its capacity. Indeed, the construction of new portions of infrastructures implies extremely high costs and may be limited by physical and political constraints. Hence, in the last decades the attention of the academic and industrial community has been more and more attracted toward the design of procedures to improve the efficiency of capacity exploitation.

Train timetables are arguably a critical element affecting the efficiency of exploitation of the existing capacity. For this reason, considerable effort has been devoted to improve the timetable design. Indeed, researchers have been working on the train timetabling problem since the fifties. Since then, an important number of

optimization approaches have been proposed in the literature, most of them based on macroscopic representations of the infrastructure. A few of these approaches have been successfully used in practice as decision support tools that allow the design of improved timetables with respect to the previous manual procedures, in terms of both feasibility and efficient capacity exploitation. In general, macroscopic approaches allow the consideration of rather large railway infrastructures thanks to the introduction of a high level of abstraction. However, the gains in the efficiency of capacity exploitation may be limited by this high level of abstraction. Indeed, assumptions on the actual capacity are made to cope with the lack of details inherent to macroscopic approaches. Microscopic approaches, instead, offer a finer level of details, which allow the modelling of various specific aspects of the railway system, such as the routes followed by the trains arriving/departing from a station, the utilization of the tracks and platforms, the actual train circulation policies applied in practice, etc. This not only allows a higher confidence on the feasibility of the designed timetable, but it also promotes a more efficient capacity exploitation. Because of their high complexity, the use of microscopic approaches is typically restricted to the main stations and junctions of the railway infrastructure. However, in the current context of scarce capacity, we deem necessary that microscopic approaches take part in every timetabling procedure performed. Moreover, we believe that the scope of microscopic approaches should be extended to larger portions of the infrastructure.

During the elaboration of timetables, planned temporary unavailabilities of portions of infrastructure due to maintenance are taken into account. However, only macroscopic aspects of the infrastructure and the maintenance activities are typically considered. Due to the absence of details of these macroscopic approaches, it may be necessary to verify the microscopic feasibility of the designed timetable before its actual implementation. In addition, unforeseen maintenance needs may emerge at any time, e.g., due to a device malfunction, and this may require a rearrangement of the established timetable. Currently, there is a limited number of approaches in the literature to tackle the problem of producing or rearranging timetables to cope with the performance of maintenance activities. In particular, we remark a complete

absence of microscopic approaches, which would ensure the feasibility of the designed timetables while optimizing the capacity exploitation and considering specific aspects of maintenance activities such as the imposition of temporary speed limitations. Nonetheless, the use of such specialized approaches is becoming increasingly important due to the heavy maintenance activities which are envisaged for the near future in many European countries. Indeed, some parts of the railway infrastructure in these countries have attained their ageing limit, as for example in France. Moreover, all the European railway networks have to be adapted to comply with recent regulation aiming to ensure their interoperability.

1.2 Objectives

The objective of this thesis is to promote the efficient railway capacity exploitation by means of producing feasible and efficient timetables. More specifically, we define two main objectives, each one containing specific sub-objectives.

Our first main objective (O1) is to **propose an approach for designing feasible timetables which ensures an efficient exploitation of the railway capacity**. We believe that the use of a microscopic approach is essential to attain this objective. Currently, the typical scope of microscopic approaches is limited to a single station or a complex junction. Here, microscopic approaches are used to validate the assumptions on the capacity of these stations and junctions made by macroscopic approaches. These assumptions may underestimate or overestimate the actual capacity of these elements, leading to its inefficient exploitation or the production of infeasible timetables, respectively. However, we believe that microscopic approaches may be used for optimizing timetables, rather than only validating the ones produced under questionable assumptions. To this end, as a first sub-objective (O1.a), we aspire to **prove that microscopic approaches can be used for portions of the infrastructure larger than one specific station or junction**. Indeed, several specific procedures and interactions within the railway system, that

can be represented in microscopic approaches, can not be modelled using macroscopic ones. They may have a strong impact on the produced timetable. Hence, as a second sub-objective (O1.b), we aspire to **find evidence of the need to use microscopic approaches**.

Our second main objective (O2) is to **propose an approach to rearrange timetables to cope with the performance of maintenance activities** of the railway infrastructure. The proposed approach has to consider all the main realistic aspects associated to the performance of infrastructure maintenance activities, such as the presence of specialized rolling stocks and the imposition of temporary speed limitations. Indeed, these specific aspects are often neglected when tackling the problem of timetable rearrangement, although we believe that their impact is far from negligible. As a first sub-objective (O2.a), coherently with what done for timetable production, we aim to **demonstrate the capability of a microscopic approach to deal with relevant-size instances** of this problem. Indeed, due to the heavy maintenance activities envisaged for the near future, there is an increasing need of approaches effectively tackling this particular problem for real circumstances. The few existing approaches are all based on macroscopic representations of the infrastructure without considering the realistic aspects of maintenance discussed above. Therefore, as a second sub-objective (O2.b), we wish to **demonstrate that the proposed approach exhibits a more efficient capacity exploitation than the ones obtained by using the existing methods** tackling this problem.

Although we believe that microscopic approaches cannot currently replace macroscopic ones, we claim that they must become a crucial complement to them. At the tactical level, when timetables are designed for the whole railway network of a country, the macroscopic approaches have to integrate the microscopic ones in an iterative way. At the pre-operational level, when timetables are to be rearranged to cope with maintenance activities or other unplanned capacity needs and the scope is typically more limited, the microscopic approaches can play a leading role in the optimization process. At this level, the macroscopic approaches may be used to validate the

microscopically rearranged timetables in wider scopes. In both cases, we think that the capability of microscopic approaches to efficiently tackle instances covering more than a single station or junction is capital. On the one hand, this may reduce the number of iterations needed to converge toward an optimal country-size timetable. On the other hand, this will allow infrastructure managers to promptly and suitably respond to unplanned capacity needs.

1.3 Contributions

The contributions of this thesis are in alignment with the objectives described in the previous section.

First, we propose RECIFE-TTP, a *Mixed-Integer Linear Programming (MILP)* formulation based on a microscopic representation of the infrastructure that aims to solve the timetabling problem. RECIFE-TTP has been developed as part of the decision support tool named RECIFE (REcherche sur la Capacité des Infrastructures FErroviaires - Research on the Capacity of Railway Infrastructures) introduced by Rodriguez et al. (2007). Thanks to the fine level of details considered in RECIFE-TTP, we are able to model specific circulation constraints that can accurately emulate the railway traffic in real circumstances. In our experiments, we tackle a real case study of the French railway network containing several stations. We solve this case study with a number of algorithms implementing variants of RECIFE-TTP, including different scheduling and routing settings and objectives functions. The results of the experiments show that RECIFE-TTP can be successfully applied for the design of timetables concerning rather large portions of infrastructure. Moreover, the results support the claim that a microscopic timetable optimization is necessary to ensure the feasibility of the produced timetables.

Second, we propose RECIFE-MAINT, a microscopic approach for timetable rearrangement to cope with maintenance activities. Indeed, we consider specific aspects of these activities such as the planning of maintenance trains and the imposition of

temporary speed limitations. To the best of our knowledge, this is the first microscopic approach in the literature to tackle this specific problem considering these aspects. We implement an algorithm that emulates the current practice to tackle this problem, which we use as a benchmark. We propose as well different algorithms based on RECIFE-MAINT, aiming to attain good computational performance. After a thorough experimental analysis using a real case study, we demonstrate the applicability and good performance of our algorithms with respect to the current practice.

1.4 Outline

The rest of this thesis is organized as follows.

In Chapter 2 we introduce several concepts and procedures characterizing a typical railway system. More specifically, we start by defining the railway capacity and the general railway service planning process. We give extended descriptions of the timetabling and maintenance planning procedures, as they are tightly related to the main contributions of this thesis. In addition, we present the actual implementation of these procedures in France, as a practical example. Finally, we thoroughly describe the two main approaches to represent the railway infrastructure: macroscopic and microscopic. We discuss the differences among them to compare their strengths and limitations when used for producing and modifying timetables.

In Chapter 3 we review the train timetabling problem in more detail. We start by introducing several particularities of this problem and offering an extended literature review of relevant scientific contributions. Then, we thoroughly present RECIFE-TTP, our microscopic formulation to solve the train timetabling problem. We introduce the algorithms we propose and a real case study based on the French railway system which is then used for our experiments. The results are presented and discussed.

In Chapter 4 we deal with the problem of timetable rearrangement to cope with maintenance activities. After reviewing the limited available literature, we present the formalization of this particular problem. Then we introduce RECIFE-MAINT, our microscopic formulation to allow timetable rearrangement in presence of maintenance activities. We perform a series of experiments tackling the case study introduced in Chapter 3, considering the performance of several maintenance activities.

Finally, in Chapter 5 we summarize the topics covered in this thesis. Then, we present our conclusions and the main research issues that will need to be explored in the future.

Chapter 2

An overview of the Railway System

2.1 Introduction

The *railway system* is a complex and large compound of fixed and movable devices, operating protocols, entities and crews that interact in a coordinated and regulated manner to allow the circulation of trains for transportation of passengers or freight. In this chapter we introduce several concepts, terminologies and procedures that are relevant for the understanding the rest of this thesis. For a comprehensive review of the railway system we refer the reader to Profillidis (2014).

The first concept that we introduce concerns the definition of the railway infrastructure capacity, or simply capacity. The efficient exploitation of this capacity can be seen as the general objective of the main contributions of this thesis. More specifically, as discussed in Section 1.2, these contributions are centred around the elaboration and modification of train timetables. Indeed, timetables are a central element of any mass transportation system, and particularly of the railway system. Train timetables, as published in stations or web sites, are aimed to be simple and easy to understand by the users of the system. Despite this intended simplicity, the

actual process of producing a timetable is a very complex and time consuming task.

The timetabling process is a part of the even larger and more complex railway service planning process. Thus, to well understand the timetabling, we first present the whole railway service planning process as it is regulated in Europe and implemented in European countries. Then, we explain how the timetabling fits in it and how it relates to the other processes involved.

The second part of the railway service planning process which we discuss in detail is the maintenance planning. This is very important for appreciating the contributions of this thesis which concerns the integration of maintenance and timetables.

Besides describing the general railway service planning, timetabling and maintenance planning processes, in each dedicated section, we present how they are actually performed in the French railway system.

In the final section of this chapter, we introduce the main elements existing in a modern railway infrastructure and the typical way to represent them. More specifically, we focus on those elements that are considered for timetabling and maintenance planning. Furthermore, we present two different approaches for representing them: macroscopic and microscopic.

To summarize, the rest of this chapter is organised as follows. Section 2.2 gives the definition and classification of railway capacity. Section 2.3 describes the railway service planning process. Sections 2.4 and 2.5 introduce timetabling and maintenance planning, respectively. Section 2.6 presents the two main approaches used to represent the railway infrastructure and discusses their differences. Finally, Section 2.7 concludes this chapter.

2.2 Railway Capacity

The study of the capacity in a railway system and its efficient exploitation has become a central concern for railway infrastructure managers. Although capacity seems to

be a self-explanatory term in common language, when used in a formal context, it is a complex term that may have numerous meanings and for which several definitions have been given. Krueger (1999) defines the capacity in railway systems as a measure of the ability to move a specific amount of traffic over a defined rail line with a given set of resources under a specific service plan. The capacity in railway systems is classed by Abril et al. (2008) in the following types:

- Theoretical capacity: It is the number of trains that could run over a segment of the infrastructure, during a specific time interval, in a strictly perfect, mathematically generated environment with the trains running permanently and ideally as close as possible between each other. Theoretical capacity ignores the effects of variations in traffic and operations that occur in reality. The theoretical capacity can be viewed as an upper limit for the line capacity since it is not possible to actually run the number of trains that can be worked out mathematically.

- Practical capacity: It is the traffic volume that can be moved on a line with a reasonable level of reliability. It is calculated under more realistic assumptions which are related to the level of expected operating quality and system reliability. It is the capacity that can permanently be provided under normal operating conditions which usually represents around 60-75% of the theoretical capacity (Kraft, 1982).

- Utilized capacity: It is the traffic volume operated over the network. It reflects actual traffic and operations that take place. It is usually lower than the practical 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 our research, we are mainly concerned about an efficient exploitation of the practical capacity, therefore, in the rest of this work we refer to it simply as *capacity*,

for which we use the definition given by Hachemane (1997): The maximum number of trains that can be operated on a given railway infrastructure, during a specific time interval, given a set of operational conditions and a given service quality.

2.3 Railway Service Planning Process

Planning of public transportation systems, and particularly the railway system, is a very complex, expensive and time consuming task. Indeed, the railway system is composed by a large number of sub-systems with very particular functions that must interact in a very coordinated manner. The interoperation of these subsystems have to be carefully planned to achieve a high level of reliability and quality of the railway service.

One of the major objectives of the planning process is the production of the *annual railway service plan*. As in several other mass transportation systems, the railway service is often defined in a yearly plan. The annual railway service plan is basically a compound of documents that specify several aspects of the service, e.g., the annual commercial timetable, the capacity allocation plan, the crew schedules, etc.

To cope with the high complexity of the railway service planning process, efforts have been made to formalize it. Typically this is done by dividing the process into a set of specific activities that may be treated separately. These activities are mainly involved in the production of specific parts of the annual railway service plan. Despite the fact that a formal process is defined in every railway system, in practice, it is not always possible to strictly follow the planning process. In the following two subsections, first we describe the nominal railway service planning process, and then we shortly review a practical case: the planning process in France.

2.3.1 Nominal Planning Process

A recent approach presented by Klabes (2010) formalizes the planning process for the case of the European railway system. Based on the latter, we present a diagram in Figure 2.1 resuming the railway planning process.

Two main types of entities are involved in the whole planning process: The *Infrastructure Manager (IM)* and the *Railway Undertaking (RU)*. According to the definitions given in CIT (2014): An IM is a body or firm responsible for establishing, managing and maintaining railway infrastructure, including traffic management, control-command and signalling. A RU is a public or private entity that provides services for the transport of goods and/or passengers by rail.

Indeed, a separation of the traditional single, state-controlled entity to oversee the whole organization of the railway system took place in the last decades in European countries. In the early nineties, it was remarked a steady decline of the railway industry with respect to the expanding road transportation system and it was decided to revitalize the former. To this end, a series of European directives were issued, starting with Directive 91/440/EEC, aiming at fostering the development of competition and the realisation of larger and more efficient markets, stimulating additional employment and increased investment. Since the first directive was issued, four Railway Packages have been published (in 2001, 2004, 2007 and 2011), which consisted of a group of directives that further specified several organizational and interoperability aspects to be addressed by European countries.

As in several other organizations and industries, the planning and managing of the railway service can be divided into three levels: Strategic, tactical and operational. Strategic level activities concern mainly traffic demand forecasting and the acquisition and renovation of resources, e.g., railway infrastructure and rolling stock, to cover the target demand. The tactical level covers the allocation of these resources, while their consumption is dealt with at the operational level. The start and end dates of each level are typically defined with respect to the date in which

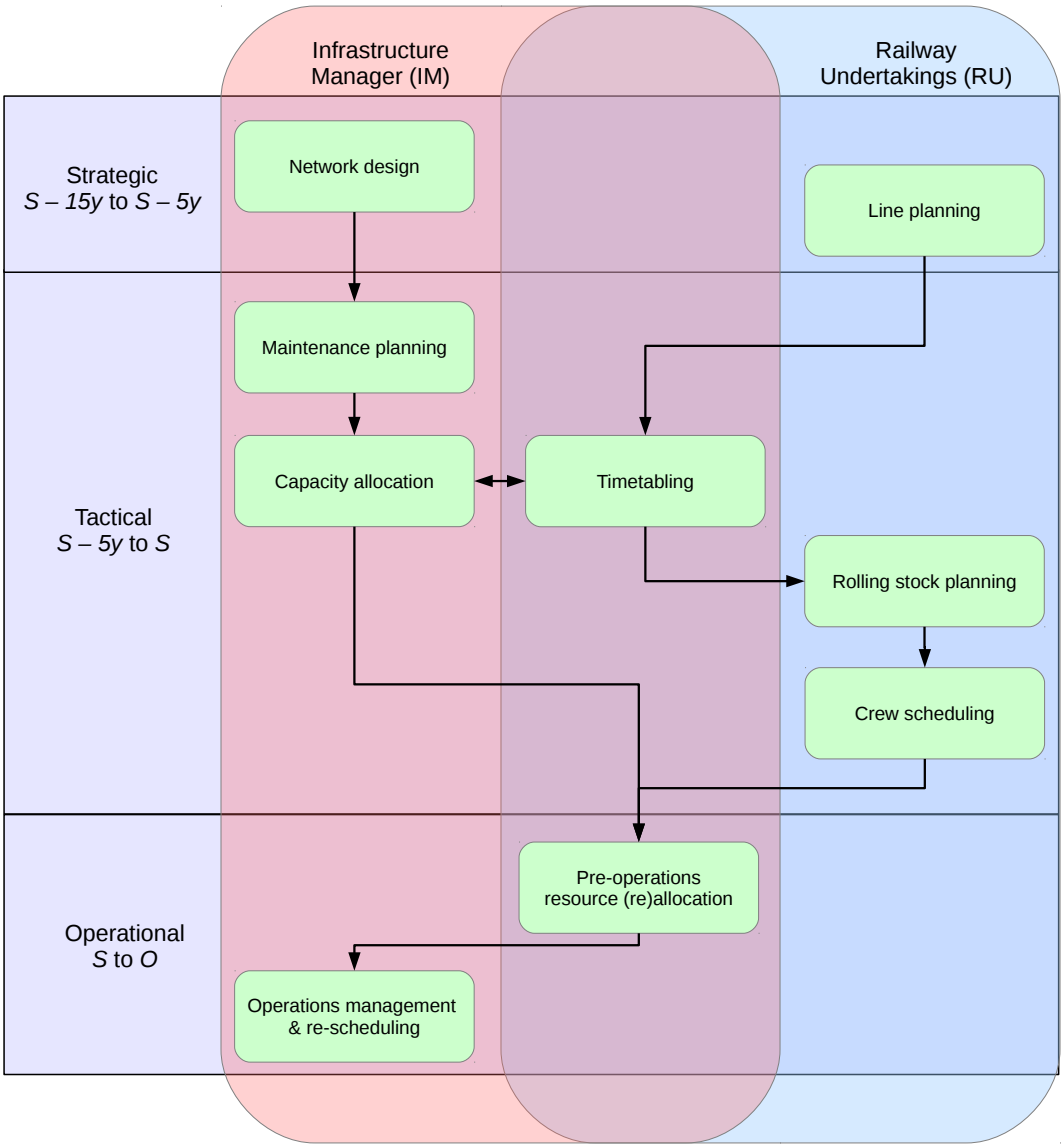


Figure 2.1: The Railway Service Planning Process. Activities are classed firstly by the type of entity responsible of their performance (IM and RU), and secondly, by the temporal level in which they are performed (strategic, tactical and operational). S: start date of the annual railway service plan, O: day of operations.

the annual railway service plan starts (S) and the day of operations (O). Figure 2.1 depicts the planning activities at the three levels.

The nominal flow of activities, represented by the directed arrows in Figure 2.1, is from top to bottom, where the activities at the same height may be carried out simultaneously. However, in practice, the flow of activities is not entirely sequential. In some cases, when a problem arises while performing some activity and it can not be solved in said activity, it may be necessary to return to the precedent one so it can be worked out. The next sub-sections give a more detailed description of the activities performed at each level of the planning process.

2.3.1.1 Strategic Level

The strategic level of planning typically goes from 15 to 5 years prior to the start date of the annual railway service plan ($S - 15y$ to $S - 5y$). It includes two main activities:

- Network design: This activity consists in the elaboration of plans for the construction or the modification of the railway infrastructure. This activity is typically carried out by the IM in cooperation with the public authorities. RUs may also participate in the process by providing studies of customer habits, demand and projections. Reasons to modify or build new infrastructure include: changes of the travel requirements, increased demand of capacity and implementation of new technologies or standards. This activity is in constant execution and revision, projects are treated individually and require a substantial amount of time to be established and approved, several years in some cases. Relevant literature addressing the network design problem can be found in Magnanti and Wong (1984), Hooghiemstra et al. (1999) and García-Archilla et al. (2013).

- Line planning: This activity consists in the definition of the train lines to be considered in the timetable design. A train line can be defined as an *itinerary* between two designated stations in the network, generally major ones, that is

to be recurrently traversed by a train. Note that the itinerary may also include a set of intermediate stations. The frequency or desired schedules of the trains and the type of the required rolling stock are also defined along with the lines. During the planning process it is quite common to consider direct connections between lines and total travel time for passengers, as these aspects are often referred as quality indicators for a line plan. This activity is mainly performed by the RUs, but it might also require some collaboration with the IM and to some extent with the regional governments. Several scientific contributions were made to tackle the line planning problem, among these: Bussieck et al. (1997), Claessens et al. (1998), Goossens et al. (2006), Schöbel (2012) and Fu et al. (2015).

2.3.1.2 Tactical Level

The tactical level of planning covers the period of 5 years before the start date of the annual railway service plan (from $S - 5y$ to S). Mainly five activities are performed at this level:

- Maintenance planning: Also known as the maintenance scheduling problem, it consists in defining a maintenance plan that meets the maintenance needs of the network. A *Maintenance Activity (MA)* can be defined as a set of actions performed for retaining a system or an item in, or restoring it to, a state in which it can perform its required function. Typically, the maintenance plan includes all preventive MAs and, in some cases, time slots reserved for possible corrective MAs. Note that the projects of modification and construction of new infrastructure, defined in the network design activity, are generally treated as MAs as well. The IM is in charge of establishing the maintenance plan which is normally done before the definition of the timetable. We will present a detailed description of railway maintenance activities and of their planning in Section 2.5.

- Timetabling: The objective of this activity is to obtain a timetable, which can

be considered as one of the most important elements of the railway transportation system. This problem has received significant attention in the literature where it is mostly known as *Train Timetabling Problem (TTP)* or train scheduling problem. Each RU establishes the desired schedules for its requested train lines, while the IM is in charge of producing the timetable for all RUs' requests. Moreover, the RUs must negotiate between them and the IM to solve possible incompatibilities. We will present an extended description of the TTP in Section 2.4. Then, we will propose a detailed review of the literature and a novel approach to solve this problem in Chapter 3.

- Capacity allocation: The purpose of this activity is to define a capacity allocation plan compatible with all trains scheduled in the timetable. This plan consists in the assignment of track routes and platforms in stations to each train while respecting their schedules. The IM is the main responsible of this activity. In case of allocation impossibility, the RUs are contacted for negotiations. In the literature, this problem is also known as track allocation problem, train routing problem, train path allocation problem and, in some cases, train platforming problem. Some relevant contributions are: Caprara et al. (2011), Lusby et al. (2011), Borndörfer et al. (2012) and Schlechte (2012).

- Rolling stock planning: Consists of finding an assignment of the rolling stock to the trains scheduled in the timetable. This may also include the scheduling of empty rides and shunting movements. Typically, it is sought to minimize of the number of vehicles, or the total cost, necessary to meet the requirements of the timetable. The RUs are the main responsible for the generation of the rolling stock plan. Scientific contributions for the rolling stock scheduling problem, also known as the vehicle scheduling problem, can be found in Abbink et al. (2004), Maróti (2006), Steinzen et al. (2010) and Cacchiani et al. (2012).

- Crew scheduling: The crew scheduling problem consists of generating crew duties to operate trains at minimal cost, while meeting all work regulations and operational requirements. The RUs are responsible for their crews, therefore

they are in charge of defining their schedule. Although the majority of scientific contributions proposing methods to solve the crew scheduling problem are destined to the airline domain, they can also be applied in the railway context. Some specific contributions aiming to solve the railway crew planning problem are Caprara et al. (1997), Abbink et al. (2005), Şahin and Yüceoğlu (2011) and Jütte and Thonemann (2012).

Despite the fact that the relations between these activities are shown in Figure 2.1 as sequential, in reality, all activities in this level are strongly inter-dependent between each other. Indeed, several authors agree that ideally all these activities should be performed as only one. However, because of the high complexity of the resulting problem, it can not be currently effectively solved, e.g., Higgins (1998), Kroon et al. (2007) and Forsgren et al. (2013).

2.3.1.3 Operational Level

This level concerns the activities that take place since the start date of the annual railway service plan until the operations day (S to O). Typically, the activities considered at this level concern the operations management and the adaptation of the annual railway service plan to cope with unanticipated resource requirements, unforeseen events and real-time disturbances. Note that timetables are particularly susceptible to these adaptations and they may have to be modified, this operation is known as re-timetabling or timetable rearrangement. We identify two main types of activities in this level:

- Pre-operations resource (re)allocation: Despite the fact that the annual railway service plan is already defined, validated and currently in use, it is very likely that new, unanticipated resource requirements may emerge. These can be for instance: new freight trains due to an exceptional convoy, unexpected corrective maintenance due to a device malfunction, changes on the crew rotations due to strikes, etc. A common characteristic of these evens is that although they are unforeseen, some time is available to deal with them, in general, up to

ten days before the operations day. Therefore, this type of activities covers the unforeseen resource (re)allocations requested between the start of the annual railway service plan and ten days before the operations day (S to $O - 10d$). The IM is responsible when the resource required to be (re)allocated is capacity, e.g., due to unforeseen MAs, new trains requests, etc. Instead, RUs are responsible if the resources are rolling stock or crews. In Chapter 4, we will introduce a novel approach to rearrange a timetable to cope with unplanned infrastructure MAs.

- Operations management & re-scheduling: The objective of this activity is to oversee, manage and coordinate the train traffic during operations. Perhaps the most critical action is the re-scheduling of train circulations. In general, the train re-scheduling problem aims to reduce the delay experienced by trains to overcome incidents or unforeseen incompatibilities, which may also include train cancellations. This activity is performed by the IM. However, when major modifications to the train schedules are required, the RUs are contacted for negotiations. Typically, the railway infrastructure is divided into control zones where re-scheduling decisions are made locally. Therefore, an important issue to consider is the impact of local decisions in the whole network. Moreover, due to the real-time nature of this problem, solutions need to be rapidly calculated in spite of possible inefficiencies. Traditionally, this activity is performed by the train dispatchers but in recent years several approaches have been proposed to automate or support re-scheduling decisions, as: D'Ariano et al. (2007), Corman et al. (2011) and Pellegrini et al. (2014). Rather different approaches have been proposed for tackling traffic disruptions in which negotiations with the RUs are necessary. Examples of the latter can be found in Dollevoet et al. (2012), Jespersen-Groth et al. (2009) and Veelenturf et al. (2011). For a detailed review of approaches dealing with the *real-time Railway Traffic Management Problem (rtRTMP)* we refer the reader to Cacchiani et al. (2014).

2.3.2 Railway service planning process in France

Although the process is quite well defined in the literature and regulations, in practice, its actual realization does not precisely follows the description given in Section 2.3.1. As an example, in this Section we depict the current railway service planning process in France. As an introduction to this, we start by briefly describing the main actors operating in this country.

Since its origin in 1938, the state-owned company *Societé Nationale des Chemins de fer Francais (SNCF)* managed the railway infrastructure and coordinated all train traffic in France. However, to comply with the new European railway regulations (railway packages, mentioned in Section 2.3.1), the SNCF experienced a series of restructuring processes. Perhaps the most important reform was the implementation of the separation of the IM from the RU. In 1997, *Réseau Ferré de France (RFF)* was created to operate as the French IM, and SNCF operates as a RU. As of January the 1st 2015, a new restructuring occurred where two entities were defined, RFF became SNCF Réseau, which acts as the IM, while SNCF Mobilités serves as the main RU in the country. Other RUs are: Rail4Chem, SNCB, CFL, Véolia Cargo, ECR, VFLI, Europorte, CFL Cargo, Colas Rail, Eurotunnel, etc.

The railway service planning process in France can be divided into five phases, as thoroughly described in RFF (2009) and RFF (2011): Strategic, Structuring/Planning, Production, Adaptation and Last-Minute/Operational. A schema depicting this planning process is shown in Figure 2.2. Even if this process is different from the one reviewed in Section 2.3.1, it still follows the main lines of the European regulations. The next subsections briefly explain the main goals and activities for each phase.

2.3.2.1 Strategic Phase

The strategic phase in France goes from 15 to 5 years before the start of the annual railway service plan ($S - 15y$ to $S - 5y$). The main objectives are to:

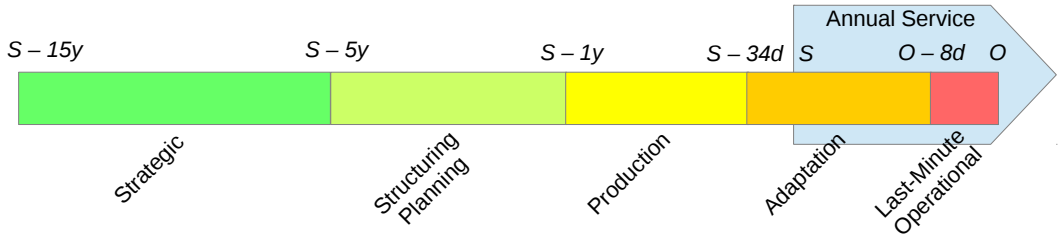


Figure 2.2: The five phases of the French railway service planning process. S: start date of the annual railway service plan, O: day of operations.

- Anticipate and estimate the traffic evolution.
- Identify and outline the infrastructure needs to match with the current and estimated traffic demand.
- Roughly define the capacity requirements and allocations for: maintenance (particularly large infrastructure renewal and modification projects) and commercial trains (high speed intercity or international services).
- Advise the RUs concerning their investments in rolling stock.

2.3.2.2 Structuring/Planning Phase

The Structuring/Planning phase starts at $S-5y$ and lasts until $S-1y$. The objectives of this phase are to:

- Produce the maintenance plan and allocate the necessary capacity.
- Estimate the time supplements necessary for train circulations to cope with the execution of maintenance in the main lines of the network.
- Produce a weekly schedule of the capacity allocations for maintenance.
- Analyse, estimate and pre-allocate the necessary capacity for recurrent commercial circulations.

2.3.2.3 Production Phase

This phase starts at $S-1y$ and lasts until 34 days before the start date of the annual railway service plan ($S-34d$). This phase has mainly one objective: to finalize the production of the annual railway service plan. This is achieved by the performance of the following activities:

- Refinement of the weekly maintenance plan, define a schedule for each MA.
- Processing of the complete lot of trains or line requests made by the RUs.
- Resolution of conflicts between requests.
- Production, validation and publication of the annual timetable which considers the capacity consumption by MAs.

2.3.2.4 Adaptation Phase

The adaptation phase starts at $S-34d$ and lasts until 8 days before the operations day ($O-8d$). The main activities in this phase are:

- Processing of late train requests and allocating the available capacity.
- Specification of the capacity allocation for maintenance. If ultimately some MAs will not be performed, the corresponding reserved capacity is liberated.
- Resolution of possible conflicts and unforeseen events.

2.3.2.5 Last-Minute/Operational Phase

This phase lasts for the 8 days prior to the operations day, including it ($O-8d$ to O). The main activities are:

- Processing of last-minute train requests following a first-come first-served policy.
- Final liberation of the non specified capacity allocation for maintenance.

- Resolution of conflicts, disturbances and other unforeseen events.

- Management of real-time operations.

2.4 Timetabling

As the main contributions of this thesis are related with timetabling generation and rearrangement, we provide in this section a more detailed description of the general timetabling process. Moreover, we review a practical example of a real case: train timetabling in France.

The process of elaborating a train timetable, is an essential activity in the railway service planning. Indeed, timetables represent a core element in the railway transportation system as they are the main references to coordinate the traffic in the system. In a railway context, train timetables specify information concerning trains circulations, i.e., schedules. Train schedules define the departure and arrival time of trains at stations and other relevant points on a railway network. This information is relevant for all the parties involved in train operations, i.e., customers, IMs and RUs.

Timetabling consists in the assignment of a schedule to each train request. Each train request, also known as *train-path request*, is defined by an itinerary and a desired schedule. The itinerary of a train defines its origin, destination and, if needed, a set of intermediate stops. The desired schedule specifies the desired times for each arrival/departure event of the itinerary.

In several railway systems, two types of timetables are typically defined: commercial and working timetables. A *commercial timetable* defines the scheduled arrival and departure times of trains at stations. It is mainly conceived to be employed by the users of the railway system to help them find their transportation plan. A *working timetable* includes timing details at every station, major junction and other significant locations in the railway network. It is designed to be used by the railway

operating staff. In our research, we focus on the production and rearrangement of working timetables, thus, unless an explicit distinction is made, in the rest of this work we refer to them simply as timetables.

An important aspect to consider in the production of timetables is that they must respect the *green wave policy*. This policy states that a train must follow its itinerary without having to stop or reduce its speed due to a *conflict* with another train. A conflict exists when two trains travelling at the planned speed would concurrently require the same track segment, which would result in the reduction of speed or even complete stop of one of the concerned trains. In other words, a train must only stop at its final destination and its intermediate stops, if any. There are several reasons to do so: better passengers comfort, lower power consumption, compliance with safety and security issues, etc.

The problem of defining a timetable is formalized as the TTP. The fist objective of the TTP is the construction of a *feasible* timetable. A timetable is feasible if all planned train circulations are conflict-free and respect the capacity and safety constraints of the railway system.

Traditionally, train timetables were designed by railway planners, which solved the TTP by hand, based on their experience. Naturally, it is difficult to guarantee the feasibility of manually obtained timetables. Moreover, they most likely present an inefficient exploitation of the capacity. Recently, the application of optimization approaches to solve the TTP is becoming an essential part of the timetabling process. In Chapter 3, we give a detailed review of the main approaches found in the literature.

Although the application of these approaches have undoubtedly improved both the feasibility of timetables and the efficiency of capacity exploitation, they do so at different degrees. The improvement degree obtained for these two aspects, particularly for the latter, highly depends on the modelling and representation of the railway infrastructure. Indeed, several ways to represent the infrastructure are used in the literature, which can be classed as macroscopic and microscopic. We review

in detail and compare these two types of representations in Section 2.6. Moreover, in Chapter 3, we present a microscopic approach for solving the TTP.

2.4.1 Timetabling in France

In this section we briefly present the train timetabling process in France described in RFF (2009) and RFF (2011). The construction of the annual timetable takes place in the Production phase of the French railway planning process, which begins one year before the start date of the annual railway service plan (S), and ends with the *schedule certification*, which is the validation and publication of the annual timetable and takes place 34 days before S. However, it is possible to modify the timetable even after the schedule certification, typically to allocate new train requests. This operation is also known as re-timetabling or timetable rearrangement and may take place in the Adaptation and Last-Minute/Operational phases of the planning process.

As soon as the Production phase starts, the RUs can submit the train-path requests ("sillons") to SNCF Réseau, the French IM. To ensure that a train will be scheduled in the timetable, RUs must submit the corresponding request at least 240 days before S. Train-path requests made within this period are called *on-time train-path requests*.

Those requests made after $S-240d$, also known as *late train-path requests*, are also considered by SNCF Réseau, but they may be only accommodated in the available capacity after the allocation of on-time train-path requests. Late requests can be submitted even after the start date of the annual railway service plan: they are accepted until 8 days before the operations day (O).

Train-path requests are still possible between $O-8d$ and the actual operations day O, these are known as *last-minute train-path requests*. Last-minute requests are allocated following a first-come first-served policy in the available capacity. Moreover, RUs must agree that these trains may be scheduled without respecting the green wave policy or a complete compliance to the desired schedule.

On-time and late train-path requests are handled by the planning unit of SNCF Réseau, and they are submitted via specially designed computer tools: *Pathfinder* for international train-path requests or *Fiche-Sillon* for the national and regional ones. These tools also perform an initial feasibility verification of the train-path request, i.e., the existence of a valid route and the operability of the desired schedule.

Last-minute train-path requests are directly transmitted to the traffic and operations management unit of SNCF Réseau which deals with them in the shortest delay possible: depending on the available capacity, the train-path request is either immediately rejected or scheduled.

Typically, the platform occupation in stations is planned jointly with the scheduling process. Moreover, the planning of shunting movements is also performed during the scheduling process. However this is done only for stations where the capacity is scarce.

Throughout the whole timetabling process, a computer based tool called DEMIURGE is used to aid planners to: assess the capacity consumption and availability, optimize the scheduling procedure and evaluate the robustness of the timetables.

2.5 Maintenance Planning

Because the execution of maintenance activities has a direct impact on the (re)scheduling of train circulations, which is in fact the main contribution of this thesis, we deem important to describe in detail which are the main maintenance operations carried out and how are they planed. Therefore, in this section we begin describing the different types of maintenance operations in a railway system. Then we describe their impact on the available railway capacity, and finally we give a practical example by explaining the maintenance planning process in France.

Maintenance is performed for retaining a system or an item in, or restoring it to, a state in which it can perform its required function. In case of the railway system,

maintenance can be divided into two main categories depending on the object of the activities: infrastructure and rolling stock. In this work, we are concerned with the planning and execution of railway infrastructure maintenance activities, therefore, we refer to them simply as Maintenance Activities (MAs).

MAs are necessary to maintain the good state of the railway infrastructure, allowing trains to circulate safely and fluidly, thus ensuring the availability of the railway capacity. Indeed, while performing MAs, the train circulations in the concerned locations are impacted. This impact depends on the type of MA performed, which in general implies circulation interdiction on some track segments and temporary speed limitations on neighbouring ones. This means that, during the performance of MAs, the available capacity is reduced. In other words, MAs are capacity consumers.

Depending on the characteristics and the timing in which MAs take place, they can be classed in the following categories:

- Inspection: It is mainly destined to measure and evaluate the current state of the infrastructure. Although most elements present in the infrastructure are periodically revised, special attention is given to the rail, overhead lines, signals and switches. Inspection MAs are typically short in duration and fairly regular. Special equipments on board of moving vehicles are used to improve the efficiency of the inspection process. In general, inspection MAs do not require a downtime as they can be performed during operations.

- Corrective: It is performed when an element of the infrastructure brakes or malfunctions. Corrective MAs are generally triggered by the occurrence of an unscheduled event. Typically this type of maintenance requires an immediate and significant downtime of the service in specific locations, which may cause important disruption on the traffic in the whole network. This type of MAs are in general very costly not only because of the originated traffic disruptions but also because they often require urgent reparations. Moreover, although several protection mechanisms are in place, the failure of an element often causes the

failure or damage to other elements in the system, which increases the costs of this type of MA, e.g., the derailment of a train caused by a faulty switch may also cause severe damages to other contiguous tracks and equipments.

- Preventive: This type of MAs consists in the adjustments, reparation or replacement of certain elements in the infrastructure to reduce the probability of failures and malfunctions, i.e., to limit the necessity of expensive corrective maintenance. Inspection activities are often considered into this category as well, however we prefer to keep them separated because inspections do not require to modify in any way the infrastructure and they have in general a much shorter duration. While the majority of preventive activities requires a downtime of the concerned location, some of them can be performed during operations. There are mainly three approaches used to define the planning of preventive MAs:
  - Manufacturer schedules: In general, manufacturers provide a recommended maintenance or replacement schedule that are applicable in most cases regardless of the particular conditions of the system, e.g., a section of rail of more than ten years should be replaced.
  - Rules and conditions: Similarly, manufacturers or domain experts determine a set of threshold values that may condition the execution of a MA, e.g., a section of rail presenting corrugations over 1.5mm requires a grinding operation.
  - Predictive models: The elaboration of complex degradation models are used to accurately estimate the time at which an element of the infrastructure should be maintained or replaced while optimizing the relation between cost and required reliability. Predictions are made taking into account several aspects besides the manufacturer specifications, e.g., local weather conditions, actual usage rate and characteristics, information obtained during inspection activities, etc.
- Construction: The activities of construction of new infrastructure, or modifi-

cation of the existing one, are also considered as MAs. Indeed, this type of activities is very expensive and often requires several years of planning and execution. Moreover, depending on the location, these MAs may cause heavy downtime periods in specific parts of the network, e.g., the construction of new platforms in a station may require a temporal but complete shut-down of said station.

In general, the cost of performing MAs is very high and it tends to increase as time passes. For example, the total railway infrastructure maintenance cost in France in 2007 was approximately 3.2 billion Euros, which represents nearly the 30% of the total budget allocated by the state to the SNCF in that year (Putallaz and Tzieropoulos, 2012).

Besides the elevated costs of MAs, it is important to remember as well that MAs are capacity consumers. Recall that in most cases, train circulations are not allowed in locations where MAs take place. This means that the cost of unexploited capacity caused by a MA should be also taken into account when estimating the total cost. For more details concerning the estimation of these costs, we refer the reader to Putallaz and Rivier (2003).

A mechanism to limit the reduction of the available capacity caused by MAs is to, when possible, re-route trains so they run on contiguous tracks. Indeed, in several railway systems, particularly in Europe, large portions of the railway infrastructure consists of two parallel rail roads, one for each direction. In such case, during the execution of a MA, trains that are planned to use the tracks concerned by the MA itself may be re-routed to temporary circulate in opposite direction on the contiguous track. To perform such action, a security mechanism is required to avoid frontal collisions between trains or deadlock situations.

On the one hand, in tracks where this kind of action is not provided for, a temporary opposite direction mechanism may be implemented. Although this is a viable option, its use is very restrained due to safety reasons. On the other hand,

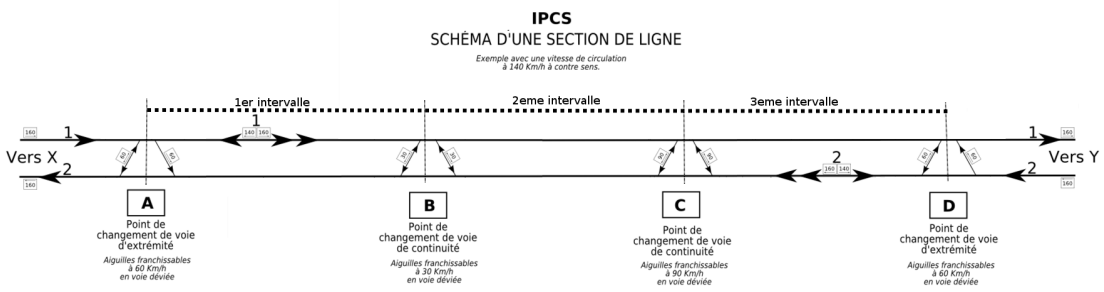


Figure 2.3: An example of an infrastructure equipped with permanent inverse-way circulation equipments (IPCSs) which allow bi-directional circulation of trains in both rail roads. This double track section connecting stations X and Y implements four track change points, A, B, C and D, which, along with the specialized signalling system, provide three IPCS intervals allowing a max speed of 160 km/h for circulations in nominal direction and 140 km/h for opposite ones. (Adapted from the Wikimedia Commons file "File:Installations Permanentes de contre sens.png" https://commons.wikimedia.org/wiki/File:Installations_Permanentes_de_contre_sens.png)

some tracks are built with a *permanent opposite direction circulation equipment*, which is also known in France as *Installations Permanentes de Contre-Sens (IPCS)* (SNCF, 1991). This equipment consists of special signalling and switching devices which provide a high level of safety for train circulations in both directions over a single track segment. Figure 2.3 shows an example of an infrastructure equipped with IPCSs.

By equipping tracks with IPCSs, the railway infrastructure becomes more flexible in terms of number of available routes for trains. This improves the recovery capabilities of a timetable when a disruption occurs and, perhaps more importantly, it reduces the capacity costs associated with the execution of MAs.

In recent years, the advantages of the use of IPCSs have been studied more in detail due to the increasing need of capacity which is also translated into the need of a more efficient exploitation of the existing one. Some examples can be found

in Putallaz (2007) and Lafont and Antoni (2012). Some European countries have recognized the benefits of these equipments and they have implemented them in their entire networks, e.g., the Sweden infrastructure supports full capacity bidirectional operation on both tracks on double-track lines (Forsgren et al., 2013).

2.5.1 Maintenance Planning in France

In this section we briefly review the current state of the railway infrastructure in France, the new maintenance policies established, and the process of planning MAs. The infrastructure maintenance strategies and policies in France have experienced important changes in recent years. After the presentation of the renowned audit by Rivier and Putallaz (2005), often simply referred as "Rivier" audit, the French IM applied a number of reforms.

Indeed, the "Rivier" audit pointed out that the state of the French railway infrastructure was rapidly degrading and most of the elements in the infrastructure had already surpassed their expected life time. This was particularly true in the segments of the network that were seldom used, i.e., track lines used by less than 20 trains per day, which represented approximately 46% of the total length of the French railway infrastructure. Moreover, the annual maintenance budget did not show any significant growth since the eighties. Instead, it was reduced (the maintenance budget in 2004 represents approximately 75% of the maintenance budget in 1981), and most of the budget was used to maintain the existing infrastructure rather than invest into renewal.

These findings, along with the projected growth of traffic demand, prompted the French government and IM to completely redesign the maintenance strategy of the infrastructure. Some of the measures implemented were: increasing the maintenance budgets, prioritizing the renewal investments, reducing the frequency but enlarging the stretch of maintenance interventions, etc. For a more detailed review of the reforms applied following the "Rivier" audit, see Putallaz and Tzieropoulos (2012).

These measures were encapsulated in a brand-new "Mass Maintenance Process", which is summarized in Goolen (2008). This new process includes the definition of three types of maintenance interventions: Preventive, Corrective and Inspection. These types of MAs are considered as *stable* because their duration and frequencies are predefined and cannot be modified. Instead, a special type of maintenance is defined only for specific maintenances such as renewal or construction of new infrastructure. These *non-stable* MAs in general require longer durations and more frequent interventions than those envisaged by stable maintenances. The different types of maintenance interventions are resumed in Table 2.1.

Category	**Type**	**Duration**	**Period**	**Weekdays**	**Frequency**
stable	Preventive	6h	night	Sun-Fri	8-10 weeks a year
		4h	day	Mon-Fri	
	Corrective	3h	day/night	Sat/Sun	once every 2 weeks
	Inspection	1h	day	Mon-Fri	3-5 days a week
non-stable	Special	6h-24h	day/night	Sun-Sat	N.A.

Table 2.1: Characteristics of the different types of maintenance interventions in the French maintenance planning process.

The period of time in which the railway capacity is reserved to perform MAs is known as *maintenance window* ("fenêtre travaux"). Typically, the entire duration of a maintenance window is reserved in early stages of the maintenance planning process, i.e., train circulations are not scheduled into these periods. If the concerned location of the infrastructure is equipped with IPCSs, the maintenance window is referred as a permeable window ("fenêtre poreuse"). Differently from regular maintenance windows, permeable windows actually allow the scheduling of trains inside them, provided that time supplements are assigned to the circulations.

The precise operations to be performed within the maintenance windows are specified in later stages of the planning process. Once these precise operations are defined, the maintenance window is transformed into several maintenance intervals

("plages travaux") which correspond to each of these specific operations. The definition of these maintenance intervals may release some of the reserved capacity, i.e., the maintenance window is typically larger than the combined maintenance intervals. In this case, the released capacity can be allocated to commercial trains.

In France, the complete process of maintenance planning, i.e., since the identification of maintenance or construction needs until the allocation of the necessary capacity and other resources, starts roughly eight years before the annual railway service plan (S) and ends at most seven days before the operations day (O).

Between $S-8y$ and $S-4y$ the strategic maintenance planning for the main lines is performed. Strategic planning for regional lines is carried out between $S-6y$ and $S-4y$. Strategic planning includes the identification, definition and preliminary sizing of the principal MAs, particularly non-stable ones, such as construction and renewal.

Between $S-4y$ and $S-2y$, all MAs (stable and non-stable) are planned on a weekly basis. To this end, planners make use of a computer based tool called PLATINE. This tool allows the creation of weekly based schedules, where maintenance windows are assigned. It also takes into account the availability of resources required for performing the MAs, such as required crews and specialized rolling stocks and other materials.

Between $S-2y$ and $S-1y$, the weekly plan is refined, i.e., the dates and times of all maintenance windows is specified. In this stage, PLATINE is also used, as it allows a refinement of the initial weekly schedule. It takes into account the impact on standard circulations in the concerned locations and attempts to reduce the costs when setting the schedule for a maintenance window.

Starting from $S-1y$ and ending at most on $O-7d$, the specific maintenance operations are defined for each maintenance window, i.e., the maintenance intervals are generated. In this stage, regular maintenance windows may release capacity which can be used to schedule new commercial trains. In case of permeable windows,

a rearrangement of already scheduled train circulations may be required. To this end, dispatchers make use of another computer based tool called SIOUCS. This tool performs rearrangements of the trains' schedules by following a batteries approach, which we review in detail in Section 4.3.1. After $O - 7d$, all maintenance windows disappear and any capacity not reserved by a maintenance interval is released, which may allow the scheduling of last-minute train requests.

2.6 Railway Infrastructure

In this section we review the main elements present in a railway infrastructure and the typical way to represent them. Indeed, the railway infrastructure consists of a large number of devices and operating procedures that have been in constant evolution since the appearance of the first railway systems. We focus on the subset of these devices and procedures that are usually taken into account for (re)timetabling procedures. For a comprehensive review of the devices and operating procedures in railway infrastructures, we refer the reader to Retiveau (1987) and Theeg et al. (2009). In general, two main approaches are used to represent the railway infrastructure: macroscopic and microscopic, which we detail next in the next sections.

2.6.1 Macroscopic representation

Macroscopic representations are based on the abstraction of large elements existing in the infrastructure, e.g., stations and tracks connecting them. These elements are conventionally expressed as a graph $G = (V, E)$ where the set of vertices V represents the stations and the set of edges E represents the tracks connecting two stations. Figure 2.4 shows a graph-based macroscopic representation of a part of the Dutch railway infrastructure, presented in Kroon et al. (2007).

Although graphs are widely used in macroscopic representations, several authors in the literature propose several variants to express certain peculiarities that are

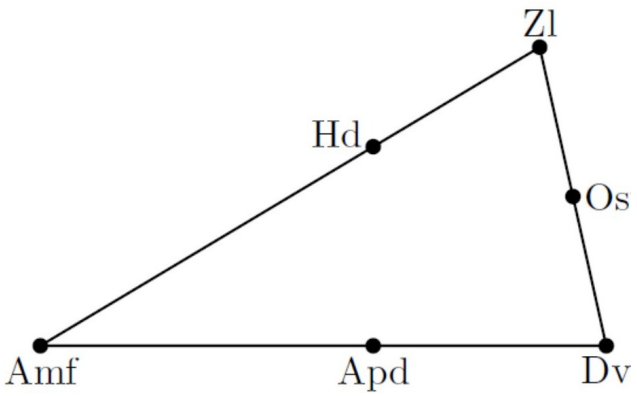


Figure 2.4: A macroscopic representation of a part of the Dutch railway infrastructure, presented in Kroon et al. (2007). It represents the triangle Amersfoort (Amf), Deventer (Dv) and Zwolle (Zl). Other relevant stations are Apeldoorn (Apd), Harderwijk (Hd) and Olst (Ost).

relevant for the purpose of their research. In some cases, a single edge connecting two vertices may be enough. However, to express more accurately the number of tracks between stations, a multi-graph may be used instead. Tracks are often considered as mono-directional, which means that trains can only circulate in one direction. To express this, directed edges may be used. Weighted graphs are also commonly used. On the one hand, the weight of the edges may express several things, as: the number of tracks (as an option to multi-graphs), the maximum number of trains that can use the track at the same time, a cost for using the track, etc. On the other hand, weighting the vertices is less frequent, and it is generally used to express their capacity, which we will discuss later in this section. Moreover, in some cases, vertices are also used to express bifurcations, sidings, or simply locations where trains may stop.

Commonly, macroscopic representations use approximations to define circulation and capacity constraints, e.g., the minimum headway time between two trains on the same track segment, the maximum number of trains that a station can host at a given time, etc.

To control the train's utilization of the capacity, it is necessary to identify the time

period in which a train is utilizing a track. Typically this is done by the definition of a *running time*. In macroscopic approaches, the running time can be interpreted as the time needed for a train to circulate from one station to another interconnected by a track. The running time depends on several factors as the type and length of the tracks, topology of the terrain, type and speed of the rolling stock, etc. Some macroscopic approaches also define minimum and maximum running times to include some flexibility on train circulations.

To ensure the safety of train circulations in a track connecting two stations it is necessary to impose a temporal or physical separation between the trains utilizing said track. Typically, in macroscopic approaches a *headway time* is defined for couples of types of rolling stock using a track. The headway time is the minimum time interval that must separate the entrance in a track of two trains using it. The calculation of the headway time is based on several characteristics of both the tracks and the rolling stock, such as maximum speed, weight and brake capabilities of the rolling stock, etc. Note that some models also define a headway time for the exit of trains from a track. Moreover, the headway time can also be applied for two trains using the same track in opposite directions, if allowed. In such case, the headway time is at least the maximum running time of the train entering the track first.

The inbound/outbound track layouts in stations or junctions determine a number of routing options for arriving/departing trains. These different routing options have an impact on the trains' minimum values for running and headway times. Since macroscopic approaches do not differentiate between these routing options, the most conservative value must be used to guarantee the absence of conflicts. Hence, the actual capacity is underestimated. In other words, running and headway times are often larger than the ones actually possible, which leads to a possibly inefficient exploitation of the capacity.

Moreover, the capacity should be also managed inside stations. Indeed, stations also are formed by a set of tracks that allow trains to pass, stop or perform manoeuvres. These tracks are also finite resources that should be managed as they impact

the capacity of a station. In macroscopic approaches stations are represented as vertices in a graph and, for sake of simplicity, the station's capacity is often disregarded, e.g., Caprara et al. (2002). In this case, the general assumption is that stations have enough tracks to accommodate the trains that are scheduled to stop or pass by them.

When obtaining a theoretically feasible timetable using a macroscopic approach that does not consider stations' capacities, in practice the timetable itself may be infeasible, e.g., consider a station with only two platforms and a timetable where four trains are scheduled to stop at said station at the same time. In macroscopic approaches that consider stations' capacities, usually they are defined as the maximum number of trains that a station may handle at a given time. More sophisticated approaches may define types of tracks, e.g., platforms, manoeuvre tracks, etc., and assign them to the trains in the timetable. However, the actual feasibility of the planned use of the tracks is typically not guaranteed.

2.6.2 Microscopic representation

In microscopic representations, a high level of details is sought and the actual elements existing in the railway infrastructure are explicitly modelled. This allows the modelling of sophisticated circulation behaviours that can actually reproduce the manner in which trains interact with each other and the railway infrastructure in the real world.

Ever since the first railway systems were developed, different methods were conceived to allow trains to circulate in a coordinated and safe manner. Currently, the most widely used method is the signalling system coupled with the route-lock sectional-release interlocking system. Next we briefly describe these two systems and the elements necessary to model them. For a comprehensive review of these and other safety systems, we refer the reader to Theeg et al. (2009).

The signalling system is used to direct traffic and keep trains clear of each other at all times. It is based on the division of the railway infrastructure into *block section*s.

A block section is a segment of track which can be used by at most one train at a time. The length of a block section is variable; it depends on several factors, e.g., topology of the terrain, braking capacities of the trains authorised to use them, etc.

The entrance and exit of a block section are guarded by entry and exit *signal*s. A signal is a semaphore that gives the train drivers some information about the utilization state of the following block section(s). Usually, signals have three colours, i.e., red, yellow and green, which are also called aspects.

In a three-aspect signalling system: The red aspect indicates that the driver must stop since the next block section is currently being utilized by another train. The yellow aspect indicates that the driver can proceed but with caution because the block section after the next one is currently being utilized by another train. Commonly, a yellow aspect implies a considerable reduction of the train speed. The green aspect denotes that the next two block sections are clear so the driver can proceed at regular speed.

To automate the signalling system, it is necessary to detect the presence of a train on the tracks. Several technologies have been developed for this purpose, e.g., track circuits, axle counters, etc. A *Track Detection Section (TDS)* is a section of track where the presence of a train can be automatically detected. A single TDS can form a block section, however, for practical reasons, it is quite common for a block section to contain several TDSs. One of these reasons is strictly related with the route-lock sectional-release interlocking system, which is explained below. Figure 2.5 shows an example of a microscopic representation of a small segment of a railway infrastructure containing the elements introduced: block sections, signals and TDSs.

The interlocking system is in fact a controlled arrangement of the signalling system that prevents train movements that may create an accident. It is designed so that it is impossible to display a proceed signal unless the route to be used by the train is proven safe. To model this behaviour, the notion of *reservation* is used. Hence, before a train can occupy a block section, the latter has to be reserved by

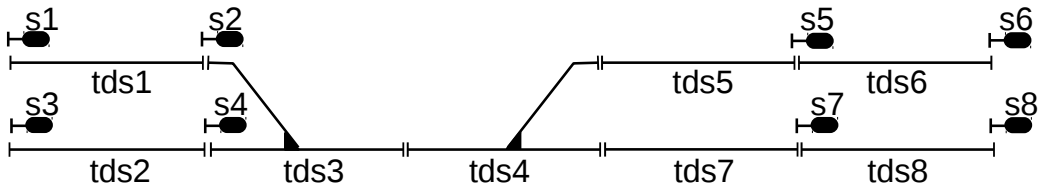


Figure 2.5: Microscopic representation of a small section of a railway infrastructure. The three main elements are represented: (I) eight TDSs: tds{1-8}. (II) eight signals: s{1-8}. (III) eight block sections: BS_s1-s2, BS_s3-s4, BS_s2-s5, BS_s2-s7, BS_s4-s5, BS_s4-s7, BS_s5-s6, BS_s7-s8.

said train. To reserve a block section, two conditions must be satisfied: First, the block section is not being occupied, or reserved by another train. Second, all the TDSs inside the block section are set and locked on the desired positions. Should the block section contain several TDSs, all of them are reserved at the same time. This part is called *route-lock*.

As the train runs inside a block section, its presence is detected on the TDSs forming it. When the train finishes the occupation of a TDS, it is released and then can be reserved by another train. The procedure of releasing independently the TDSs forming a block section is known as *sectional-release*. As a seldom deployed alternative, in the case of *route-release*, the release of all TDSs in the block section is synchronized, i.e., it occurs at the same time, when the last TDS is released. The *sectional-release* is particularly useful in locations where TDSs are shared by multiple block sections, as it helps reducing the separation between trains and, thus, maximizing the exploitation of the capacity, e.g., in large an busy stations.

Figure 2.6 shows an example of the route-lock sectional-release system. Two trains (t1 and t2) circulate in the section of the infrastructure presented in Figure 2.5. The set of coloured boxes constitutes the so called blocking time stairway, see Hansen and Pachl (2014). Note first that block sections BS_s2-s5 and BS_s4-s7, used by trains t1 and t2 respectively, share two TDSs: tds3 and tds4. t1 precedes t2, and it reserves

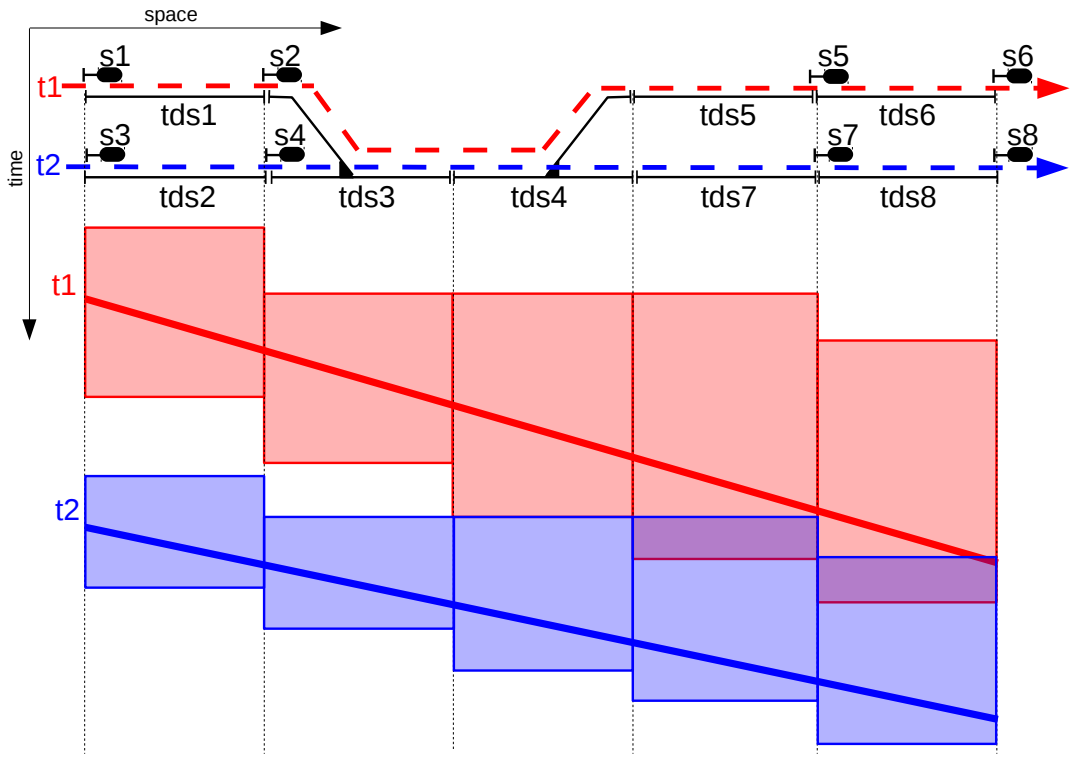


Figure 2.6: An example of the route-lock sectional-release system. Two trains circulate in the section of the infrastructure presented in Figure 2.5: t1, using BS_s1-s2, BS_s2-s5 and BS_s5-s6; and t2 using BS_s3-s4, BS_s4-s7 and BS_s7-s8. The coloured boxes represents the utilization of the TDSs by t1 and t2 following a route-lock sectional-release interlocking system. Note that tds3 and tds4 belong to two different block sections that are utilized by both trains (BS_s2-s5 by t1 and BS_s4-s7 by t2). As t1 enters first, t2 may reserve BS_s4-s7 only after tds4 is released by t1, even if t1 is still utilizing tds5. Therefore, the utilization boxes of tds5 and tds7 by t1 and t2, respectively, may overlap because both TDSs belong to the different block sections (BS_s2-s5 and BS_s4-s7, respectively). The same principle explains the box overlapping for tds6 and tds8.

all TDSs in BS_s2-s5 at the same time (route-lock). Train t2 can not reserve BS_s4-s7 as long as all its TDSs are not available. Note that since the TDSs in BS_s2-s5 are independently released by t1 (sectional-release), t2 is able to reserve BS_s4-s7 as soon as tds4 is released by t1, even if t1 is still utilizing tds5. This kind of mechanics can

only be modelled using microscopic representations, and it allows a high accuracy in the scheduling process and a highly efficient exploitation of the capacity.

Let a *control zone* be the portion of the infrastructure considered in the microscopic representation. Note that the limits of a control zone do not necessarily correspond to a station. A train is the movement of a pre-determined type of rolling stock over the infrastructure to fulfil a required itinerary between an origin and a destination. It is possible that the origin (destination) of a train's itinerary is located outside the control zone. In such case, is the first (last) TDS which needs to be crossed by the train within the control zone that is considered as the origin (destination) of the itinerary. The *travel time* of a train is defined as the time elapsed between the entrance and the exit of the train in and from the control zone, i.e., the time needed for the train to fulfil its itinerary.

A train's itinerary may require to perform intermediate stops at predefined locations, e.g., stations. We call these locations *control point*s. Note that a minimum *dwell time* at these control points is often imposed. In general, the minimum dwell time defines the minimum time that a train must remain in a platform to allow passengers to disembark, transfer and board the train.

Depending on the control zone layout and the rolling stock, a train can follow different routes to accomplish its itinerary. A *route* is a complete sequence of TDSs which can be traversed by a train to fulfil its itinerary: it goes from its origin to its destination and passes by its control points, if any.

Concerning train circulations, the *occupation time* of a train over a TDS is the time elapsed between the entrance of the head of the train into the TDS and the exit of its rear. The occupation time can be divided into two parts: the running and the clearing times. On the one hand, the running time is the time elapsed between the entrance of the head of the train in the TDS and its exit. On the other hand, the *clearing time* is the time elapsed between the exit of the head of the train and the exit of its rear.

The running and the clearing times are calculated by taking into account the characteristics of the tracks (length, gradients, curves, interlocking systems, speed constraints, etc.), the rolling stock (max speed, acceleration, braking capabilities, etc.) and the routes (origin, intermediate stops, destination). When an intermediate stop is to be performed in a TDS, the running time of said TDS does not include the dwell time. This is done because the dwell time can be later adjusted for (re)scheduling purposes. Instead, the suitable deceleration and acceleration are considered in the calculation of the running time. The same sequence of TDSs can be used along several routes if different rolling stocks are employed or if different intermediate stops are to be performed.

As previously introduced in this section, due to the interlocking system, a train must reserve all TDSs contained in a block section before occupying them. The moment in which the reservation of a TDS begins is determined by the signalling system existing in the control zone. To model the interlocking system, each block section has a reference *reservation TDS* (rTDS). All TDSs belonging to a block section have the same rTDS. The *reservation time* of a TDS starts when the head of the train enters the rTDS of the TDS and lasts until the train starts occupying the TDS. In a signalling system of n aspects, the rTDS is the first TDS of the $n-1$ preceding block section. Moreover, before a train can reserve a block section, a *formation time* is necessary to set and lock all TDSs inside the block section in the desired position, plus an additional time to take into account the signal visibility distance. Analogously, once a train has cleared a TDS, a *release time* is necessary to set the TDS in its default position.

To successfully coordinate train movements while respecting capacity and safety constraints it is imperative to ensure that two trains do not utilize a given TDS at the same time. The *utilization time* of a TDS by a train is defined as the sum of the formation, reservation, occupation and release times. Figure 2.7 illustrates the different parts of the utilization time. In a conflict-free schedule, the utilization times of a TDS by two trains cannot overlap.

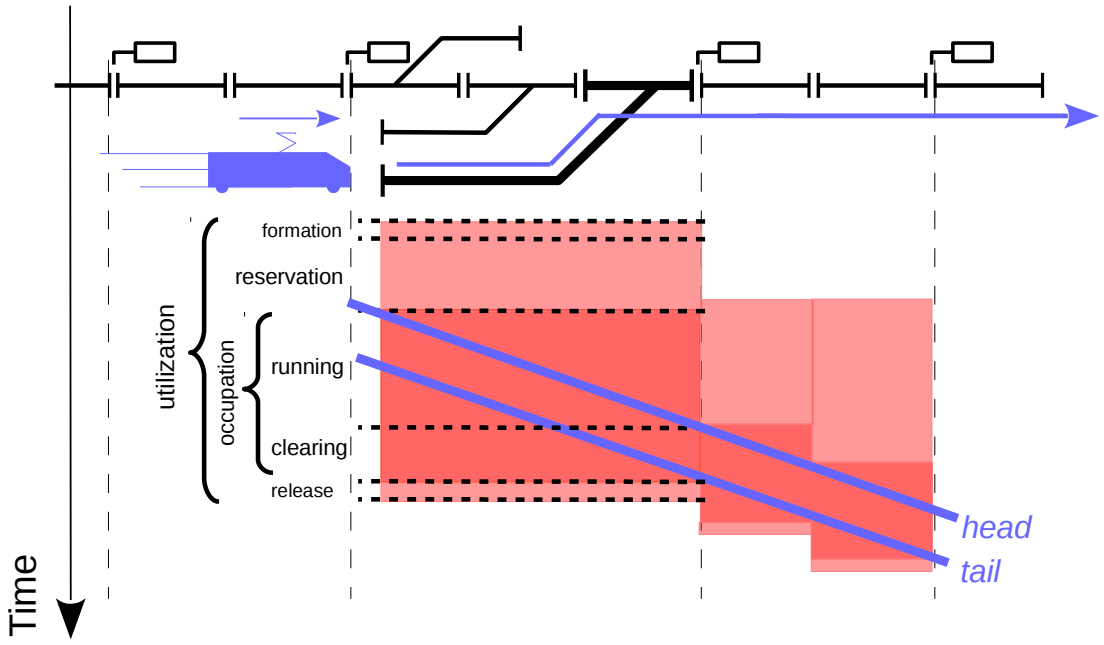
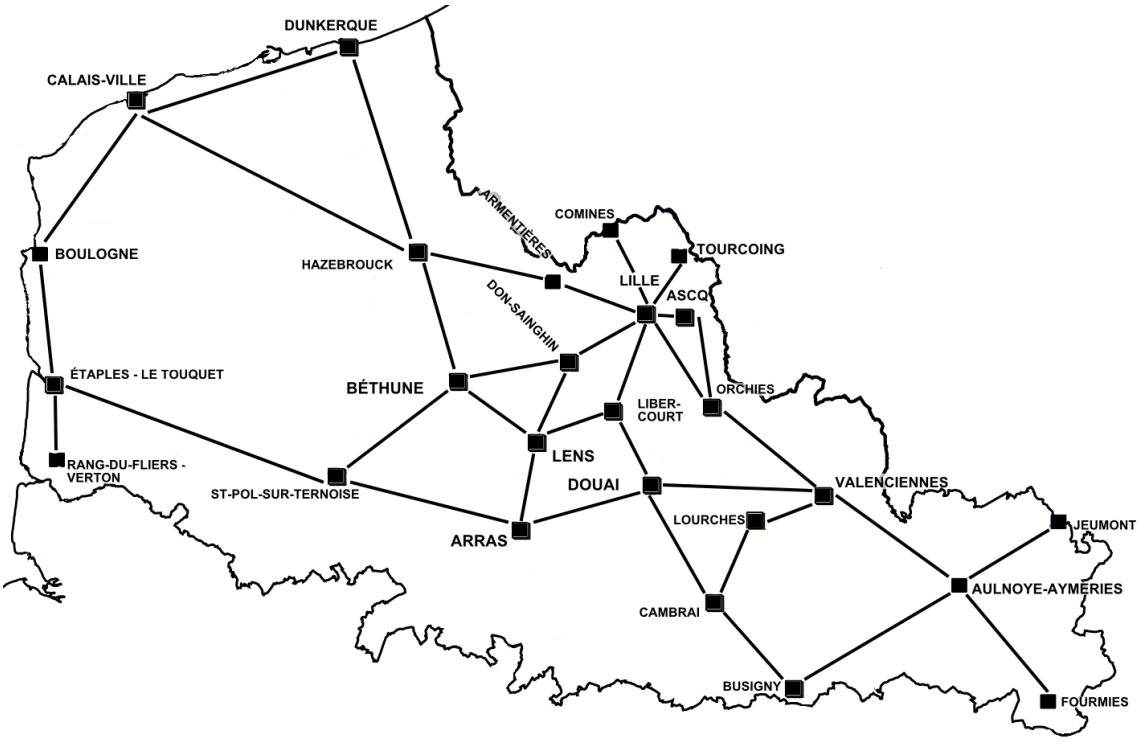


Figure 2.7: The utilization time of a train traversing a TDS in a railway infrastructure equipped with a 3 aspects signalling system.

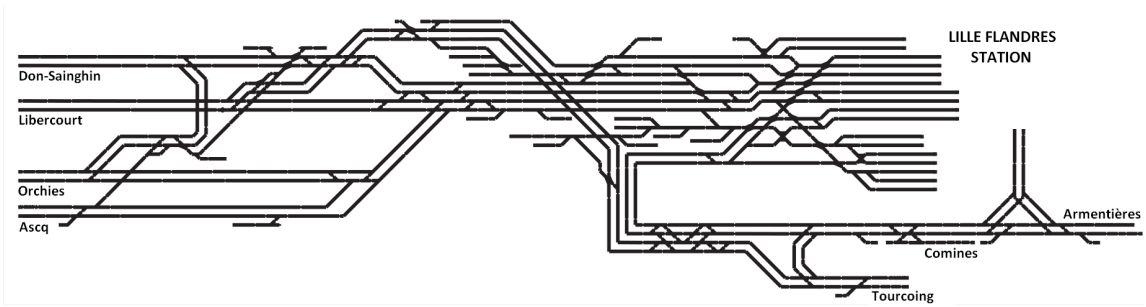
2.6.3 Macroscopic vs Microscopic representations

Macroscopic representations are a very natural abstraction of the railway infrastructure. It is quite easy to visualize the network as a set of stations interconnected by track lines. Note as well that the rail network is generally represented using a macroscopic approach in the plans designed to help the passengers find their destinations and itineraries (Figure 2.8a). Moreover, it is also very natural to conceive train circulations as the periods of time needed by a train to go from one station to another.

Instead, visual microscopic representation are indeed technical plans that, because of being highly detailed, may be hard to read and understand (Figure 2.8b). Furthermore, these visual micro-representations usually include a text-based technical specification which contains a very large amount of data that requires a thorough knowledge of railway systems.



(a)



(b)

Figure 2.8: Comparison between macroscopic and microscopic representations of the infrastructure. In (a) a macroscopic representation of the regional railway network of the Nord-pas-de-Calais, France (length: 1120 km, 26 stations and 36 double-track segments). In (b) a microscopic representation of Lille Flandres terminal station, the main railway station in the Nord-pas-de-Calais, France (length: 12 km, 299 TDSs and 734 block sections).

Indeed, the amount of data necessary to model an infrastructure using a microscopic approach is substantially larger than the one necessary for a macroscopic

representation. When microscopic elements are modelled, such as TDSs, a specific dataset for these is also required, e.g., length, allowed speed, gradients, different running times for each type of rolling stock, etc. Recall that a TDS may be very small and a track connecting two stations may easily contain hundreds of TDSs. Instead, in macroscopic representations, an equivalent dataset is only required for the whole track between these two stations. To illustrate this great difference, observe in Figure 2.8 the two infrastructure representations. Note that the micro-representation of Lille-Flandres station shown in Figure 2.8b, actually corresponds to only one node (Lille) of the macroscopic representation of the regional railway network of the Nord-pas-de-Calais (France) shown in Figure 2.8a.

It is indeed considerably easier to collect and handle macroscopic data rather than microscopic ones. To model the infrastructure, only high level data concerning the larger elements in the network are required, i.e., stations and their interconnections, which is often available for anyone. Macroscopic data concerning train circulations, such as running and headway times, is slightly more difficult to obtain. Although in principle this data can be accurately measured or calculated, time estimations are frequently used. Instead, the technical data needed to microscopically model an infrastructure is in general restricted to public access. Additionally, in several railway networks a large amount of this data may not be entirely digitalized, thus increasing the effort needed to generate a microscopic model. To obtain microscopic data for train circulations, which is likely to be non-existent in the current databases of the IMs, extensive calculations are needed. Simulation tools are typically used for this purpose.

When used to solve the TTP and related problems, the high level of details available in microscopic representations allows the definition of very elaborated models which can express very accurately the behaviour of train circulations in the railway infrastructure. On the contrary, the lack of details in macroscopic approaches allow them to model only broad aspects of the railway infrastructure. Because of this, it is quite common in macroscopic approaches to underestimate or overestimate the

actual capacity of the infrastructure, e.g., running and headway times are defined longer or shorter than they actually are along specific routes, which may lead to an inefficient exploitation of the capacity or the infeasibility of the designed timetables, respectively. Indeed, we may say that microscopic models offer a higher confidence in terms of feasibility and promote a more efficient exploitation of the capacity.

Because of the great difference in the amount of data to process and the circulation constraints to model, macroscopic approaches are in general smaller and easier to express than their counterparts. This may also imply that macroscopic approaches require less computational time to be solved. Considering these two factors, in principle, larger instances of the TTP and related problems can be tackled when using macroscopic approaches.

After reviewing the strengths and drawbacks of both representations it becomes clearer that they are rather complementary. On the one hand, macroscopic approaches are more effective to rapidly obtain solutions of large instances of problems as the TTP, although they are based on assumptions which may not allow an efficient capacity exploitation. On the other hand, microscopic approaches excel at reproducing the actual behaviour of train circulations allowing great precision of the schedules, virtually optimizing the capacity exploitation. It is important to note that the high computational time typically required by microscopic approaches, which is their most important drawback, is indeed due to a technological limitation. While macroscopic approaches will always be limited by their high level of abstractions, we can think that technological advances may at some point allow microscopic ones to overcome their drawback. In this sense, microscopic approaches have an advantage over macroscopic ones. However, in the current state of the technology, a cooperation between these approaches is necessary to tackle network-wide instances of the TTP and related problems while ensuring the feasibility and efficient capacity exploitation of the results.

2.7 Conclusion

In this chapter we introduced the main concepts, elements and planning procedures in a typical railway system. We particularly focused on those related to the production of train timetables and maintenance of the infrastructure. Moreover, we presented a practical case of such procedures to illustrate the differences between a theoretical approach and a real-world case.

The description of the general railway service planning process helps to understand in which part of this complex process are situated the main contributions of this thesis, i.e., train timetabling and timetable rearrangement to cope with maintenance activities. Moreover, the introduction of the general train timetabling procedure, coupled with the discussion of the different approaches to represent the railway infrastructure, serves as a base for Chapter 3 in which we present a microscopic approach to solve the TTP. Similarly, the description of the different types of maintenance activities and how they are planned helps to realize the relevance of the problem presented in Chapter 4, for which we propose a microscopic approach that performs an efficient timetable rearrangement to cope with maintenance activities.

Chapter 3

Train Timetabling

3.1 Introduction

As introduced in Chapter 2.4, the TTP is a critical activity of the railway service planning process. It is performed in the tactical level, more precisely after the line planning activity and before the rolling stock planning activity. The objective of the TTP is to produce a feasible timetable that responds to the commercial requirements of the customers, which are mainly, but not limited to, passenger and freight RUs. These needs are generally translated into a line schema which specifies the trains lines that are to be included in a timetable.

A timetable is feasible if all train circulations contained in it are conflict-free and respect a set of capacity and safety constraints bounded by the railway system. More specifically, these constraints are determined by the characteristics of the railway infrastructure and the rolling stock which is envisaged to be used. Moreover, other commercial requirements and quality aspects should also be considered in the timetable, e.g., the number of passengers connections at some stations. Albeit the observance of these requirements is not mandatory, the not compliance often leads to economic costs or reduction of the overall quality of the service. Furthermore, some of the commercial requirements may be incompatible with each other or with the

maintenance plan of the infrastructure. These incompatibilities should be rapidly identified, since, in most of the cases, a negotiation between the RUs and the IM is required to complete the timetabling process.

Moreover, it is desired for a timetable to be robust. The robustness can be defined as the ability of a timetable to withstand design errors, parameter variations, and changing operational conditions. A common method to increase the robustness is the addition of time supplements or time *buffer*s to the run, dwell and turnaround times of the trains in the timetable. Indeed, it is not possible to expect that the theoretical minimum values of these times are precisely respected in practice, e.g., train drivers have different conduct peculiarities, meteorological conditions may affect the running conditions, passengers take more time to disembark a train due to a door malfunction, etc. Such events may cause delay on trains during operations. The time buffers allow these delays to be contained, i.e., the initial delay is not propagated, so that some time after the event, the schedules can automatically get back to normal.

Furthermore, depending on the required type of traffic and the topology of the network, timetables can be designed to be non-periodic or periodic: On the one hand, non-periodic (or a-cyclic) timetables define the train schedules driven by market demand, e.g., train circulations are mostly requested individually, according to the needs of the customers. This may result in periods with very low or high traffic density, e.g., several trains arriving at a main station early in the morning. On the other hand, periodic (or cyclic) timetables define a train schedule for one time period (or cycle), e.g., of one hour, which is then repeated to cover the complete period of interest, e.g., of one day. Some specific considerations need to be taken into account for periodic timetables: First, the state of the network at the end of the cycle must match the state of the network at the start of the cycle. Second, the final timetable should be adjusted to reflect the traffic needs during the different times of a day, e.g., to increase the frequency of trains during peak hours. Figure 3.1 shows a space time diagram representation of a non-periodic and a periodic timetables, (a) and (b), respectively. The infrastructure consists of four stations S1 to S4 connected by double

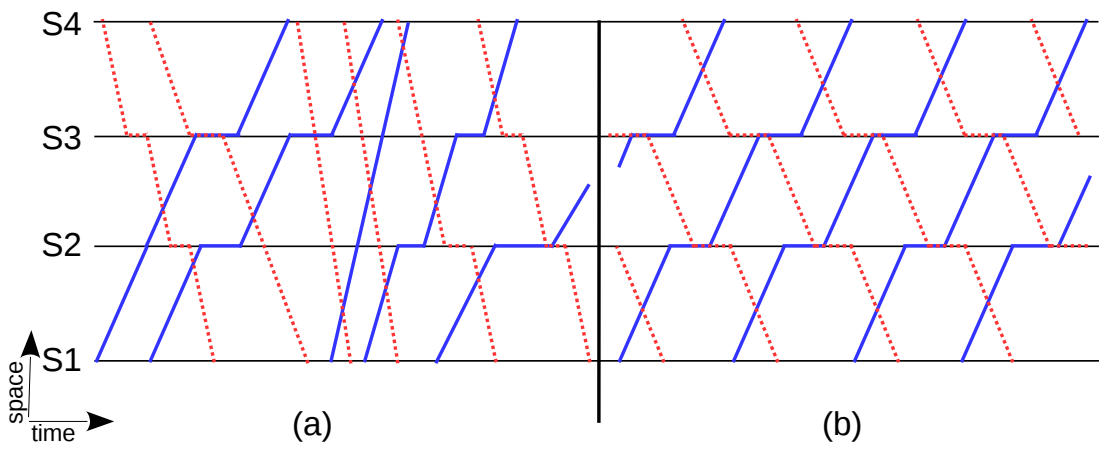


Figure 3.1: Space time diagram representation of (a) a non-periodic timetable and (b) a periodic timetable.

track segments. Note that non-periodic timetables allow more flexibility in the trains' circulations than in periodic timetables. Indeed, the intended regularity of periodic timetables require an homogenization of the train circulations, i.e., trains running in the same direction have similar speed. The advantage is that more connections may be possible in periodic timetables. It is important to note as well that the topology of the infrastructure play an important role in the design of a periodic timetable. We focus our research on non-periodic timetables, hence in the following and for sake of simplicity, we refer to the non-periodic TTP simply as TTP. The periodic TTP, instead, will continue to be referred as such.

Taking into account all the discussed requirements and considerations to manage, the preparation of a yearly timetable is a complex, costly and time consuming process that usually takes several months to be completed. The combination of all the requirements, constraints and size of the instances, which may contain a large number of trains to schedule, makes this problem very difficult to solve. Indeed, the TTP is proven to be NP-Hard (Caprara et al., 2002).

Traditionally, the timetabling process was manually carried out by the planning experts. Researchers have been interested in the TTP since the fifties but it was

only in the nineties that the benefits of using optimization approaches to solve this problem have been recognized and applied. The first two optimization approaches actually applied in practice for completely designing a timetable were: The production of the 2005 Berlin's subway timetable, presented in Liebchen (2008), and the construction of the 2007 timetable for the entire Netherlands railway network, presented in Kroon et al. (2009). More details concerning these and other approaches are presented later in this chapter. A large number of these approaches are based on macroscopic representations of the infrastructure. There are several reasons to use macroscopic approaches: the low level of detail necessary to model the railway infrastructure, the capability to deal with large size railway infrastructures within a reasonable computation time, etc. Despite the high popularity of macroscopic approaches for solving the TTP, there are some contributions based on microscopic representations of the infrastructure. Microscopic approaches not only guarantee the feasibility of the produced timetable but also excel in promoting an efficient exploitation of the railway capacity. This is possible thanks to the ability of microscopic approaches to extremely accurately represent the actual behaviour of trains and their interactions. Although microscopic approaches have gained more popularity in recent years, their usage in practice is very limited and mostly restricted to some specific parts of the infrastructure, e.g., busy stations or complex junctions. More details concerning the characteristics, strengths, limitations and differences between macroscopic and microscopic representations of the infrastructure are discussed in Section 2.6.

Currently, most railway systems around the world experience an increasing demand for railway capacity. To supply this demand, either new infrastructure must be constructed or a more efficient exploitation of the existing one must be achieved. Undoubtedly, the use of optimization tools, in particular those based on microscopic representations of the infrastructure, constitute an effective approach to promote this efficiency. Indeed, we deem that a microscopic approach to optimize the capacity exploitation must be an essential part of current and future practice in the timetabling process.

In this chapter we present RECIFE-TTP, a MILP formulation based on a microscopic representation of the infrastructure to tackle the TTP. RECIFE-TTP has been developed as part of the decision support tool named RECIFE (REcherche sur la Capacité des Infrastructures FErroviaires - Research on the Capacity of Railway Infrastructures) introduced by Rodriguez et al. (2007). More specifically, RECIFE-TTP is based on the RECIFE-MILP formulation (Pellegrini et al., 2014), which is a mathematical formulation to solve the real-time railway traffic management problem (rtRTMP). The differences and novel features of RECIFE-TTP are discussed later in this chapter.

Furthermore, we present a series of variants to RECIFE-TTP by using different objectives functions and modes for routing and scheduling. Each variant is implemented in an algorithm that is used to solve a series of instances based on a case study of the French railway network. We verify the applicability of the results obtained by RECIFE-TTP and compare the performance of the different algorithms.

The rest of this chapter is organised as follows: Section 3.2 reviews the related scientific literature. Section 3.3 formally describes the TTP. Section 3.4 thoroughly describes REFICE-TTP, our microscopic approach to solve the TTP. Section 3.5 presents the algorithms we propose. Section 3.6 first describes the real world case study which is used to perform our experiments, and then presents and discusses the results. Finally, Section 3.7 enlists our conclusions and discuss the research issues to be explored in the future.

3.2 Literature Review

Several authors have developed different techniques to address the TTP, a comprehensive review of these works can be found in Cordeau et al. (1998), Lusby et al. (2011) and Cacchiani and Toth (2012). In this section, we will briefly discuss some of the most relevant works that deal with this topic.

Although several authors make a clear distinction between periodic and non-periodic timetabling in their contributions, other authors do not specify the type of timetables they produce and simply propose a method for solving a TTP. A reason for this is that the periodic TTP can be interpreted as a constrained version of the non-periodic TTP. Note that most of the solution approaches reviewed are based on macroscopic representations of the infrastructure. For this reason we only highlight the contributions where a microscopic infrastructure representation is used.

In the next two subsections we chronologically review: First, the contributions tackling the non-periodic TTP including those which do not specifically refer to the periodic TTP. Second, the contributions for the periodic TTP.

3.2.1 Non-periodic TTP

Historically, one of the earliest optimization approaches was presented by Charnes and Miller (1956). The authors tackled the track allocation problem, which can be interpreted as the TTP in the sense that the tracks are allocated to trains following a schedule (or timetable) that is also defined by their formulation. They used a set covering formulation which they solved using a procedure similar to what today is known as column generation.

Another pioneer contribution was presented by Szpigel (1972), where the single track train scheduling problem is modelled as a job shop scheduling problem. To solve his formulation, the author used a branch and bound algorithm. This work set the basis for applying the scheduling theory to the planning of train circulations in the railway infrastructure: Train circulations are represented as a sets of jobs that must be completed sequentially. An infrastructural resource must be assigned to a job so it can be executed. Disjunctive constraints are defined to ensure the separation between trains. Other temporal and sequential constraints may be defined to comply with the specifications of the problem, e.g., desired schedules, connections between trains, etc. This modelling approach is compatible with both macroscopic and microscopic

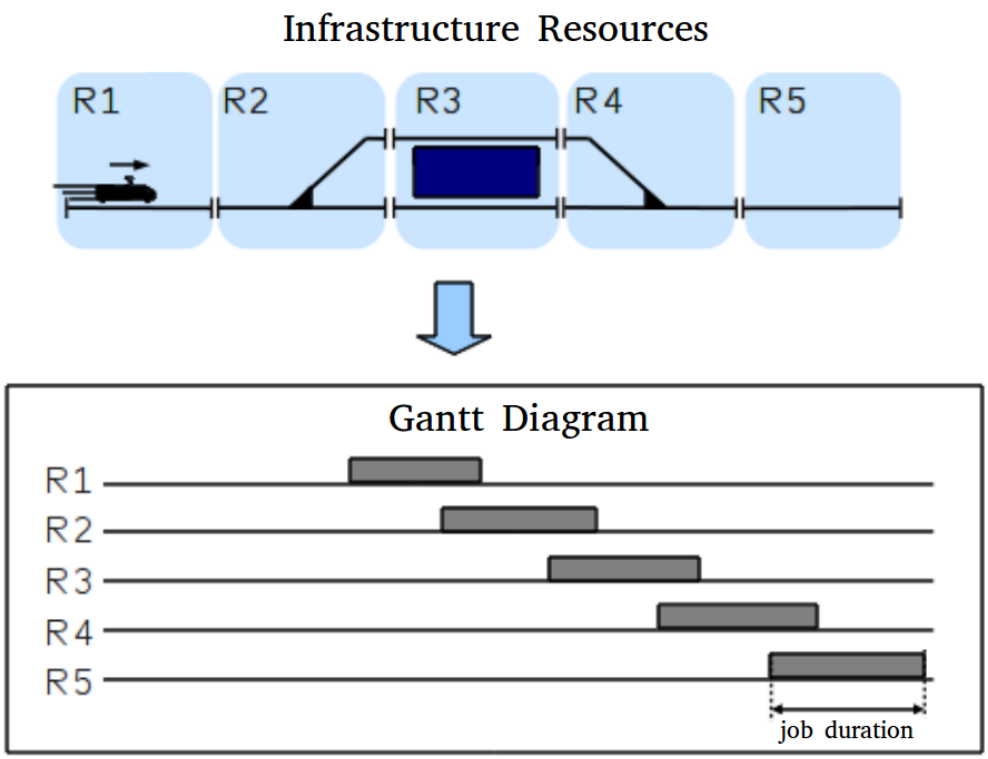


Figure 3.2: An example of the circulation of a train modelled as a job shop scheduling problem in a microscopic representation of the infrastructure.

representations of the infrastructure. Figure 3.2 shows an example of the circulation of a train modelled as a job shop scheduling problem in a microscopic representation of the infrastructure. Note that resource R3 represents a platform in a station and it allows the accommodation of 2 trains, i.e., the capacity of this resource is 2. Instead, resources R1, R2, R4 and R5 are unitary resources: only one train can use them at any given time.

In the works of Fukumori (1980) and Fukumori et al. (1987), the authors propose an algorithm to solve the TTP for a high-density railway service with multiple classes of trains. This algorithm is based on constraint propagation techniques. In their work, the authors introduced several principles for defining resource and scheduling constraints that inspired several approaches to solve the TTP using *Constraint Programming (CP)*.

Higgins et al. (1996) proposed a mathematical programming model to schedule trains on a single track line, which was applied to the Australian railway network. The authors assigned priorities to each train, which then were considered in the seek for the optimal solution through a branch and bound procedure.

In a subsequent work, Higgins et al. (1997) tackled the single line train scheduling by implementing and comparing the results of different heuristics methods: local search heuristic with an improved neighbourhood structure, genetic algorithms, tabu search and two hybrid algorithms.

Oliveira and Smith (2000) modelled the single-track railway scheduling problem as a special case of the job shop scheduling problem, by considering the train trips as jobs to be scheduled on tracks regarded as resources. Then they showed how to model the problem with CP and used Ilog Scheduler to solve it.

Caprara et al. (2002) concentrated on the problem of a single one-way track linking two major stations, with a number of intermediate stations in between. They used a directed multi-graph in which nodes corresponded to departures and arrivals at certain station at given time instants. Then, they used this formulation to derive an integer linear programming model which was solved with Lagrangean relaxation. The objective was to maximize the sum of the profits of the scheduled trains. The profit achieved for each train depended on its shift and stretch. The shift was defined as the absolute difference between the departure times from a given station in the ideal and actual timetables. The stretch was defined as the non-negative difference between the running times in the actual and ideal timetables.

Kwan and Mistry (2003) used a co-evolutionary *Genetic Algorithm (GA)* for the automatic generation of train timetables. The objectives were: first, to allocate as many train requests as possible, and, second, to discover the conflicts to be solved between the RUs and the IM. The authors used three types of populations, evolved one after another. The combination of three individuals, i.e., one of each population, resulted in a timetable. Within each step of the evolution, several timetables

were generated and evaluated. Once the termination condition was achieved, the algorithm was stopped and the best timetable was given as result.

Ingolotti et al. (2006) proposed a heuristic-based method that computed a timetable optimizing a multi-criteria objective function in a reasonable computational time. In their formulation, the authors minimized the total running time and the time deviations with respect to the desired schedules. Their method was based on the sequential assignment of a schedule for each train until the achievement of a valid timetable containing all trains.

Caprara et al. (2006) extended the formulation presented in Caprara et al. (2002) to consider several aspects of real-world TTPs. In this sense, new constraints were included to: take into account the stations capacities, consider maintenance plans in some parts of the infrastructure, manage the presence of trains with a fixed schedule and cope with outdated manual blocking systems that may still exist in some railway infrastructures.

Zhou and Zhong (2007) proposed a generalized resource-constrained project scheduling formulation where track segments and station's capacities were viewed as limited resources. The formulation was then solved using a branch and bound algorithm.

Cacchiani et al. (2008) proposed a novel formulation of the TTP. While in classic formulations each departure/arrival event is associated to a variable, in this approach, the entire schedule of a train is associated to a binary decision variable. The exponentially large number of possible schedules for a single train are dealt with a column generation approach. This approach led to an easier and faster relaxation, which allowed the authors to rapidly obtain optimized timetables of real scenarios containing between 17 and 102 stations and 40 to 221 trains.

In the work of Fischer et al. (2008), a mathematical formulation was presented and solved using a MILP solver. To improve the performance when dealing with very large instances of the TTP, the authors used a Lagrangian relaxation combined with a cutting plane approach.

Tormos et al. (2008) focused on the generation of robust timetables. The authors presented a two-step approach: First, a timetable was generated using an improved version of the method presented in Ingolotti et al. (2006). Second, using a GA, they added time supplements to the running and headway times of the timetable produced in the first step.

Wong et al. (2008) presented a MILP formulation for generating synchronized timetables. These timetables assured train connections while minimizing delays for passengers. The authors used a heuristic method to fix many of the integer variables of their formulation which was then solved using a MILP solver.

Caimi et al. (2011) focused on the problem of generating conflict-free train schedules on a microscopic model of the railway infrastructure. The authors developed a conflict graph model to describe all timing and routing alternatives for each train movement in a tree structure. Then this structure was used to identify the conflict cliques to prevent simultaneous blocking of the track resources.

Schlechte et al. (2011) presented a bottom-up approach. It started from a microscopic level where the running and headway times were computed to be later used in a macroscopic formulation. In the macroscopic level, the timetable was obtained by solving the optimal track allocation problem. Finally, the obtained timetable was transferred to a microscopic level to check its feasibility.

A recent work of Bešinović et al. (2016) proposed a hierarchical framework for timetable design which combined a microscopic and a macroscopic model of the infrastructure. The macroscopic model was used for computing a robust timetable to be afterwards converted and thoroughly analysed for solving conflicts and capacity issues at the microscopic level. The iterative micro-macro process automatically terminated once the timetable was also microscopically feasible and stable.

3.2.2 Periodic TTP

The main model used to tackle the periodic TTP is based on the *Periodic Event Scheduling Problem (PESP)*. Introduced by Serafini and Ukovich (1989), the PESP is used for the scheduling of periodic activities. The PESP is proven to be an exceptionally rich model that can be used in many applications, e.g., airline scheduling, road traffic light scheduling and of course, train scheduling.

As an example of a small PESP formulation consider three process $P1$, $P2$ and $P3$ to be scheduled in a time period T of 60 minutes, i.e., each process is to be repeated once every hour. The following propositions must be respected: $P1$ lasts between 2 and 4 minutes. $P2$ lasts between 51 and 54 minutes and it must start immediately after $P1$ finishes. $P3$ lasts between 2 and 6 minutes and it must start immediately after $P2$ finishes. A minimum separation of 2 minutes between the end of $P3$ and the start of $P1$ must be observed. These propositions can be expressed as the following set of constraints, which can be illustrated by means of a graph, as shown in Figure 3.3.

$$2 \leq (e_2 - e_1) + Tq_{1,2} \leq 4$$

$$51 \leq (e_3 - e_2) + Tq_{2,3} \leq 54$$

$$2 \leq (e_4 - e_3) + Tq_{3,4} \leq 6$$

$$2 \leq (e_1 - e_4) + Tq_{4,1} \leq 58$$

Where:

$T \equiv$ the time period, in this example 60 minutes

$e_1, e_2 \equiv$ events representing the start and end time of $P1$

$e_2, e_3 \equiv$ events representing the start and end time of $P2$

$e_3, e_4 \equiv$ events representing the start and end time of $P3$

$q_{x,y} \equiv$ a binary variable which is 0 if events e_x and e_y occurs in the same time period T, or 1, otherwise.

In Nachtigall and Voget (1996), the authors focused on the calculation of timetables with minimal waiting time for passengers changing trains. To achieve this,

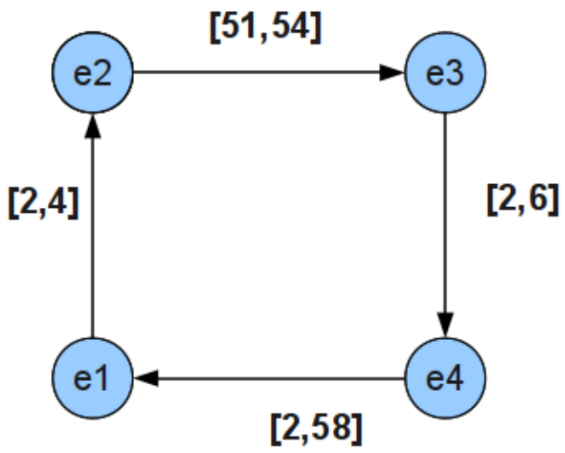


Figure 3.3: A graph representation of a small PESP formulation.

the authors presented a mathematical formulation that was solved using a two-step heuristic approach. First, an initial timetable was obtained using a greedy algorithm based on the manual procedure employed by timetable designers. Then the initial solution was improved by means of a GA.

Odijk (1996) proposed the PCG (PESP Cut Generation) algorithm to generate timetables for train operations. He did so by using a mathematical model considering periodic time window constraints as input. His model was based on the principle that arrival and departure times could be related pairwise on a clock.

In the work of Chang et al. (2000), the authors introduced a multi-objective programming model for the optimal allocation of passenger train services on an intercity high-speed railway line without branches. The objective function was formulated to minimize the total operating costs and the passengers' total travel time loss. There is travel time loss for a passenger if the train stops at an intermediate station at which the passenger does not board or disembark. The authors used a fuzzy mathematical programming approach to solve their formulation.

The work of Liebchen (2008) was based on the PESP. Two optimization criteria were considered: The maximization of the number of connections between trains and the minimization of the necessary rolling stock required by the timetable. The authors proposed two heuristics methods to reduce the computing time necessary

to solve large instances of the problem. These techniques were successfully applied to generate an optimized timetable for the subway system in Berlin. The results were obtained using a graph model of 38 nodes and 221 edges. After the application of the proposed optimization approach, the new timetable increased the number of connections from 44 to 55 and reduced by 1 the total number of vehicles required with respect to the previous timetable used in the subway system.

Kroon et al. (2009) solved the periodic TTP while considering capacity at stations. To do so, the authors presented a bi-modular approach: The first module, CADANS, was based on a macroscopic representation of the infrastructure and generated a feasible timetable for the entire railway network. The second module, STATIONS, was based in a microscopic approach and routed the trains inside the stations verifying the feasibility of the schedules produced by the macroscopic approach. CADANS consisted of a CP formulation based on the PESP, while STATIONS was based on a mathematical formulation that was solved with CPLEX. The size of the PESP formulation of CADANS included approximatively 8400 events and 7000 constraints. The authors' approach has been successfully applied for the construction of the timetable of the entire Netherlands railway infrastructure.

Liebchen et al. (2010) focused on the construction of delay resistant periodic timetables by presenting a novel combination of the traditional approach to solve the periodic TTP and a sophisticated delay management objective. The delay resistance of the generated timetables was assessed in several delay scenarios to which optimum delay management was applied.

In the work presented by Siebert and Goerigk (2013), the authors extended the scope of the PESP to include passenger paths and line frequencies. The extended formulation of the PESP, called Origin Destination aware PESP (ODPESP) was solved in two steps. An initial timetable was found for the classic PESP formulation using a CP solver, then a heuristic method was used for finding a new timetable respecting the additional constraints concerning passenger paths and line frequencies.

In Kümmling et al. (2013), the authors presented TAKT, a state of the art computer based tool to solve the periodic TTP. This tool implemented several approaches based on the PESP and other MILP formulations. Moreover, the authors encoded the PESP to propositional logic so that highly efficient SAT solvers could be used. A comprehensive review on SAT solvers can be found in Gomes et al. (2008). As the aim of TAKT was to deal with real size periodic TTPs, the authors included several strategies for decomposing large instances of the problem.

Gattermann et al. (2016) focused on the integrated problem of periodic TTP and passengers routing. To this end, the authors based their mathematical formulation on the PESP, converting it into a satisfiability problem which was then solved with a SAT solver. Experiments were performed in the long-distance German passenger railway network.

3.3 Problem Description

In this section, we describe the general characteristics of the TTP. Recall that the TTP is a part of the railway service planning process, as introduced in Chapter 2.4. It receives then as input a set of train-path requests, or the lines plan, and the output is a feasible timetable which defines a schedule for each train planned.

Each train-path request has an assigned itinerary associated with a desired schedule. The itinerary defines the origin, destination and intermediary stops, if any, of the requested train. The desired schedule defines the desired departure time for the origin, the desired arrival time for the destination and the desired arrival, departure and dwell times for each intermediate stop of the train's itinerary. Moreover, the train-path requests may also define a desired route for the train to follow. The objective of the TTP is to produce a feasible timetable containing the maximum number of requested trains while attempting to maintain their schedule as close as possible to their desired one.

On the one hand, the feasibility of a timetable is guaranteed by the observance of a set of capacity and safety constraints. These constraints are often referred as hard constraints because they must always be respected. Hard constraints are mainly related to:

- Capacity: Only one train can utilize a block section at a time. This implies also that trains can only be overtaken in locations where another track is available, e.g., stations or passing loops.

- Safety: A minimum temporal or physical space between trains must be always respected. Trains have poor braking capabilities because they are very heavy vehicles circulating, in some cases at high speed, on tracks with which they have a low grip. Because of this, the space between two trains using the same tracks must be sufficient to allow the second train to completely stop before encountering the first train if it, for some reason, comes to a complete stop.

On the other hand, the set of train-path requests are often treated as soft constraints. Soft constraints are not usually treated as mandatory, but their transgression often leads to economic costs or diminution of the quality of the timetable. Typical soft constraints are:

- Trains: In principle, all trains requested should be included in the timetable, but because of the finite nature of the capacity, this is not always possible. Trains can be associated with an economic value or profit, which can be used as a planning priority criterion. In some cases a train must necessarily be planned in the timetable, hence, the train-path request is treated as mandatory. Note that for a train to be planned in the timetable, its itinerary must be respected unless some intermediate stops are explicitly declared as optional.

- Key times: Defined by the desired schedule of the train-path request, they consist in a set of arrival/departure times. A penalty is frequently established for trains not complying with the key times. It is important to note that the key times should be at least feasible taking into account the characteristics of

the infrastructure and the rolling stock, e.g., it is not possible for a train using a rolling stock with a maximum speed of 160 km/h to go from one station to another situated 200 km away in less than 1 hour and 15 minutes.

- Time windows: Very similar to the key times, but instead of defining a precise time they define a time window in which the arrival/departure at/from a station may take place. In general, if a train cannot respect its defined time windows it is not planned in the timetable.

- Connections: Connections between passenger trains are often required at some stations. A minimum connection time to allow passengers to transfer from one train to another is defined depending on the station's size and layout.

The manner in which hard and soft constraints are modelled depends greatly on the chosen approach to represent the infrastructure. As the main contributions of this thesis are based on microscopic representations of the infrastructure, we thoroughly discuss the representation of these constraints in our microscopic formulation of the problem (Section 3.4). However, in previous works, we modelled an solved the periodic and non-periodic TTP using macroscopic approaches. For a detailed description of these approaches, we refer the reader to Arenas et al. (2013), Arenas et al. (2014) and Arenas et al. (2015).

To summarize, we formalize the TTP as follows.
Given:

- A railway infrastructure.

- A set of train-path requests.

Find a feasible timetable where:

- The maximum number of trains are planned.

- All capacity and safety constraints are respected.

- The time deviations with respect to the desired schedules are minimized.

3.4 Problem Formulation

In this section, we present RECIFE-TTP, a MILP formulation for the microscopic TTP. The goal of RECIFE-TTP is to produce a microscopically feasible timetable that minimizes the time differences with respect to desired schedules defined in the set of train-path requests, while promoting an efficient railway capacity exploitation.

There are two main inputs for RECIFE-TTP: a microscopic representation of the infrastructure and a set of train-path requests.

Concerning the infrastructure representation, all elements and abstractions detailed in Section 2.6.2 are present in our formulation, i.e., TDSs, block sections, signals and the route-lock sectional-release interlocking system, etc. Note that the precise running and clearing times are calculated taking into account the characteristics of the tracks, the rolling stock and the routes. Recall that because of the high level of details considered in microscopic approaches, the complete railway infrastructure of a railway system can not be treated in reasonable amounts of time. Therefore, as in other microscopic approaches, RECIFE-TTP focuses on specific control zones of the infrastructure. Recall from Section 2.6.2 that a control zone may include several stations and the limits of the control zone do not necessarily correspond to a station.

Concerning the set of train-path requests, they can be either directly defined by the RUs or obtained through a macroscopic approach. Indeed, the timetable obtained through a macroscopic approach can be interpreted as a set of train-path requests. Independently of the origin of the train-path requests, RECIFE-TTP designs a timetable that minimizes the time differences with respect to the desired schedules defined by them.

To generate the timetable, two main actions are performed by RECIFE-TTP: train routing and scheduling. Note that if the set of train-path requests are indeed the result of a macroscopic approach, the accurate terms would be re-routing and

rescheduling because a route and a schedule are already defined for the trains. However, for simplicity, in the rest of this section we refer to them simply as routing and scheduling, respectively.

On the one hand, routing consists in assigning to a train a route, which can be different from the one that may be defined as the desired route of the train-path request. Note that the set of alternative routes for each train is known *a priori*. In other words, routing consists in choosing the most suitable route for a train among its alternative ones. Remark that a given schedule defined for a specific route may be infeasible for other ones. Hence, when a different route is chosen, it may require a modification of the schedule as well.

Scheduling, on the other hand, consists in defining the different departure and arrival times at the control points of the train's itinerary.

Further trains may be added in the operational level (Section 2.3.1), as we will see in Section 4.3 for maintenance trains. To distinguish a train that is planned in the tactical level, i.e., during timetabling, we call it *Operational Train (OT)*. Recall from Section 2.4 that the scheduling of these trains must observe the green wave policy.

Because OTs are scheduled by RECIFE-TTP while observing the green wave policy, their predefined running times may not be modified during scheduling. Since the running times are fixed, a train's arrival time to a control point strictly depends on the departure time of said train from the previous control point. Therefore, scheduling can be viewed as the definition of the departure times from the different control points of the train. Considering the latter, we define two scheduling modes in RECIFE-TTP:

- FixDwel: The dwell times defined in the set of train-path requests can not be modified. By fixing these values, the complete structure of the trains's schedule is maintained and it can only be advanced or postponed at the beginning of its itinerary. Despite of not being very flexible, this scheduling mode is sometimes

desired by the RUs. Moreover, recall from Section 3.1 that a common practice to improve the robustness of a timetable is the insertion of time buffers to the dwell times. By not modifying the dwell times of the set of train-path requests, any contained buffer is also maintained in the solutions produced.

- VarDwell: The dwell times can be set as it is best convenient, provided that the minimum dwell times are respected. Differently from the previous mode, this one allows a high scheduling flexibility at the cost of possibly eliminating the dwell time buffers that may have been included in the set of train-path requests. Indeed, this scheduling mode observes only the minimum dwell times. However, remark that if the minimum dwell times are defined by considering buffers, these will indeed be conserved. Note that a typical timetabling issue is to find a good trade-off between the length of the dwell times and the total travel time. Indeed, by increasing the the dwell times, the robustness of the timetable may raise, but the travel time is increased as well, which is often undesirable for both RUs and passengers. Increasing the actually necessary dwell times may also decrease the capacity exploitation efficiency.

In Figure 3.4, we show an example to illustrate the differences between the two scheduling modes. Figure 3.4a shows the the desired schedule defined in the train-path request of a train departing from S1 and arriving to S4. It has two intermediate stops at S2 and S3. Two new different schedules of the FixDwell scheduling mode are shown in Figure 3.4b. Since the dwell times can not be modified, if the train is scheduled to depart from S1 "x" minutes in advance, it arrives to S4 "x" minutes in advance too (red dashed line). Analogously, if it departs "y" minutes after, it arrives "y" minutes after as well (orange dotted line), i.e., the structure of the train schedule is not changed. In Figure 3.4c, two new possible schedules of the VarDwell mode are presented. In this case, the dwell time can be changed as is best convenient, therefore, even if the train is scheduled to depart from S1 "x" minutes in advance, it may arrive "x'" minutes after to S4 due to an extended dwell time in S2 (red dashed line). Inversely, if the dwell times are reduced to the minimum allowed values, it

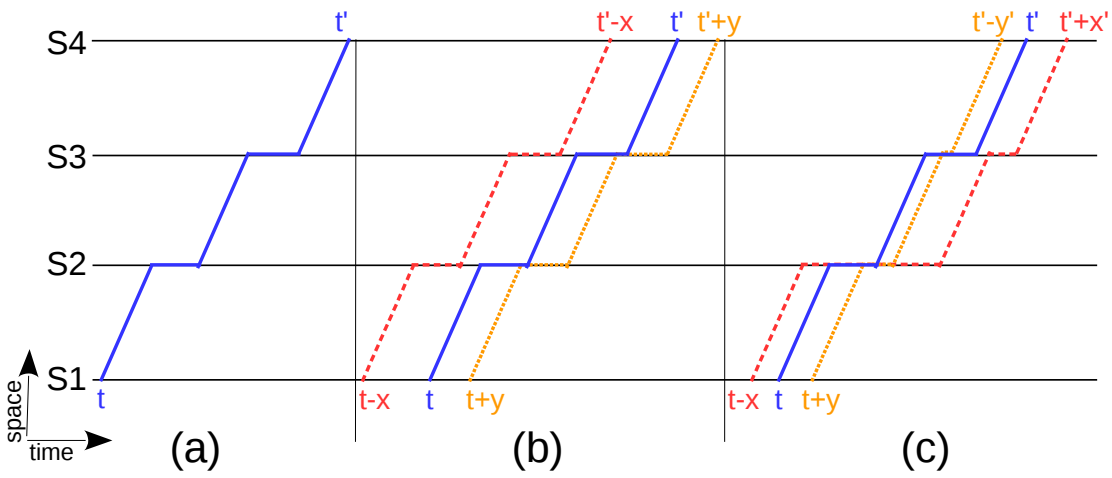


Figure 3.4: Space time diagram example of the the two scheduling modes of RECIFE-TTP. (a) The desired schedule defined in the train-path request. (b) Two possible new schedules of the FixDwell mode. (c) Two possible new schedules of the VarDwell mode.

is possible for the train to depart from S1 "y" minutes later and arrive to S4 "y'" minutes before its desired arrival time (orange dotted line).

The routing and scheduling activities are concurrently performed by RECIFE-TTP while minimizing the time differences with respect to the desired schedules defined in the set of train-path requests. To minimize these time differences, we propose two approaches: "inner" and "outer", which focus on the inside and the outside of the treated control zone, respectively.

On the one hand, the "inner" approach focuses on producing a timetable of the control zone as close as possible to the desired schedules defined by the train-path requests. On the other hand, the "outer" approach focuses on producing a timetable of the control zone that minimizes the impact produced on the adjacent control zones. Indeed, several changes to the desired schedules defined in the set of train-path requests may be necessary to ensure the microscopic feasibility and efficient capacity exploitation of the produced timetable. While the "inner" approach aims to maintain the schedules of trains as close as possible to the desired ones, it does

this at a local level, i.e., inside the control zone. However, scheduling trains that come from and/or go to adjacent control zone(s) may cause an undesired impact on the timetables of said control zone(s). This impact is explicitly minimized in the "outer" approach.

As discussed in Section 3.1, RECIFE-TTP has been developed as part of the decision support tool named RECIFE. In particular, RECIFE-TTP is based on RECIFE-MILP, introduced by Pellegrini et al. (2014), which is a mathematical formulation based on a microscopic representation of the infrastructure to solve the rtRTMP.

Recall from Secion 2.3.1 that the rtRTMP is situated in the operational level of the railway service planning process. Essentially, the rtRTMP consists in rescheduling and re-routing trains in real-time to cope with operational perturbations. The main objective of the rtRTMP is to restore the disrupted timetable to the nominal one while minimizing the trains' delays. This may imply that some trains perform unscheduled stops, e.g., in front of a signal protecting a switch, to allow the passage of another train; trains are not required to comply with the green wave policy. As mentioned, the rtRTMP takes place in the operational level, where the timetables are already published and know by the users of the railway system. Therefore, trains are not allowed to be advanced in the rtRTMP, as this might cause further problems in the system, e.g., passengers missing their trains. For a complete description of the rtRTMP and RECIFE-MILP, we refer the reader to Pellegrini et al. (2014) and Pellegrini et al. (2015).

The novel features of RECIFE-TTP include the implementation of the green wave policy and the possibility to advance the schedules of trains. The implementation of these two features involve the addition, adaptation and suppression of a number of constraints and the modification of the objective function with respect to RECIFE-MILP. In particular, the fact that trains can now be advanced may imply a strong increase in the number of symmetric solutions with equivalent objective function value. The experimental results (Section 3.6) will show that this possible increase does not compromise the computational performance of RECIFE-TTP.

In the rest of this section, we detail RECIFE-TTP: First, we introduce some assumptions we made for defining the notation we use. Then, we detail all the required input data. Finally, the MILP is presented in the following order: variables, objective functions and constraints.

3.4.1 Assumptions for the MILP notation

The notation is based on the following formalism: each element is indicated with a one-letter symbol, possibly with some superscripts and/or subscripts.

The **superscripts** aim to indicate the specific context to which the symbol refers. The superscripts used are listed in the following, with their respective meaning:
T: the symbol refers to trains,
O: the symbol refers to operational trains (OTs),
R: the symbol refers to train routes,
B: the symbol refers to rolling stock re-utilization,
α, Ω: the symbol refers to origin and destination, respectively,
P, N: the symbol refers to the previous and the following element, respectively,
V: the symbol refers to reservation,
C: the symbol refers to occupation, and
S: the symbol refers to control points.

The **subscripts** of a symbol correspond to the variation index. So, for example, Υ_t^T will be a set (of TDSs) referring to a train (superscript T) and indexed on a specific one (subscript t), while Υ_r^R will be a set (of TDSs) referring to a route (superscript R) and indexed on a specific one (subscript r).

3.4.2 Data

$R \equiv$ set of all available routes;
$\Upsilon_r^R \equiv$ set of TDSs composing route $r \in R$;
$\varphi_r^\alpha, \varphi_r^\Omega \equiv$ first and last TDSs of route $r \in R$;

$\varphi^P_{r,\upsilon}, \varphi^N_{r,\upsilon} \equiv$ previous and next TDSs w.r.t. TDS υ along route $r \in R$;

$\dddot{\Upsilon}^R_{r,\upsilon,\upsilon'} \equiv$ set of TDSs between TDSs υ and υ' along route $r \in R$;

$\Upsilon^\alpha_r, \Upsilon^\Omega_r \equiv$ set of TDSs in the first and last block sections of route $r \in R$;

$\underline{b}_{r,\upsilon}, \overline{b}_{r,\upsilon} \equiv$ formation and release time of the block section which contains TDS υ along route $r \in R$;

$T \equiv$ set of trains;

$T^O \equiv$ set of OTs, required to respect the green wave policy ($T^O \subseteq T$);

$\iota_t \equiv$ index of train $t \in T$;

$R_t \equiv$ set of alternative routes of train $t \in T$;

$\Upsilon^T_t \equiv$ set of TDSs available for train $t \in T$ ($\bigcup_{r \in R_t} \Upsilon^R_r$);

$\varphi^V_{r,\upsilon} \equiv$ reference TDS for the reservation of TDS υ (rTDS) along route $r \in \bigcup_{t \in T} R_t$, depending on the interlocking system;

$\Upsilon^C_{t,r,\upsilon} \equiv$ TDSs occupied by train $t \in T$ along route $r \in R_t$ when the head of t is at the end of TDS υ ($\emptyset$ if the rolling stock used for t is shorter than υ), a graphical representation is shown in Figure 3.5;

$\underline{e}_t, \overline{e}_t \equiv$ desired entrance time of train $t \in T^O$ and desired exit time, as defined in its train-path request;

$S_t \equiv$ set of control points for train $t \in T^O$, including its destination;

$\Upsilon^S_{t,s} \equiv$ set of TDSs belonging to control point $s \in S_t$ where train $t \in T^O$ has a scheduled stop;

$a_{t,s}, d_{t,s} \equiv$ scheduled arrival and departure times for train $t \in T^O$ at control point $s \in S_t$;

$p_{t,s} \equiv$ minimum dwell time for train $t \in T^O$ at control point $s \in S_t$;

$w^{\underline{d}}_{t,s}, w^{\overline{d}}_{t,s} \equiv$ weight of one time unit of an early/late departure of train $t \in T^O$ from control point $s \in S_t$;

$w^{\underline{a}}_{t,s}, w^{\overline{a}}_{t,s} \equiv$ weight of one time unit of an early/late arrival of train $t \in T^O$ at control point $s \in S_t$;

$w^{\underline{\alpha}}_t, w^{\overline{\alpha}}_t \equiv$ weight of one time unit of an early/late entrance time of train $t \in T^O$;

$w^{\underline{\Omega}}_t, w^{\overline{\Omega}}_t \equiv$ weight of one time unit of an early/late exit time of train $t \in T^O$;

$n_{t,r,\upsilon} \equiv$ running time of train $t \in T$ along route $r \in R_t$ on TDS $\upsilon \in \Upsilon^R_r$;

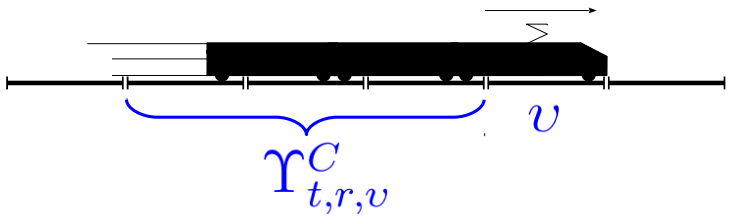


Figure 3.5: A graphical representation of $\Upsilon^C_{t,r,\upsilon}$, i.e., the set of TDSs occupied by train $t \in T$ along route $r \in R_t$ when the head of t is at the end of TDS υ.

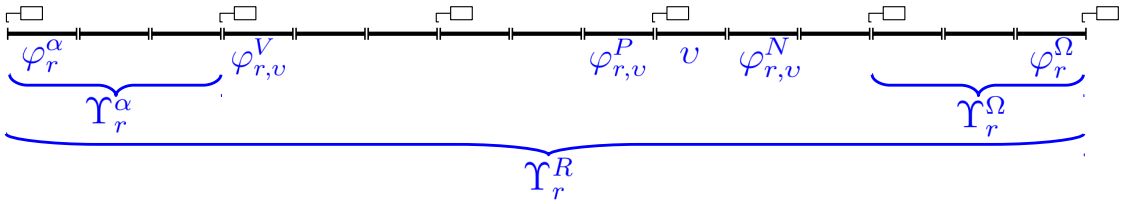


Figure 3.6: Main data concerning a TDS υ belonging to route r defined for an infrastructure equipped with a signalling system of 3 aspects.

$c_{t,r,\upsilon} \equiv$ clearing time of train $t \in T$ along route $r \in R_t$ on TDS $\upsilon \in \Upsilon^R_r$;
$T^B_t \equiv$ set of trains that result from the turnaround, join or split of train $t \in T^O$;
$\widehat{p}_{t,t'} \equiv$ minimum separation time between the arrival time of train t and the departure time of train t', provided that $t' \in T^B_t$ $(t \prec t')$;
$M \equiv$ large constant.

Figure 3.6 displays a graphical representation of the main data concerning the TDSs belonging to route r defined for an infrastructure equipped with a signalling system of 3 aspects.

3.4.3 Variables

Non-negative continuous variables:
for all triplets of $t \in T$, $r \in R_t$ and $\upsilon \in \Upsilon^R_r$:
$o_{t,r,\upsilon}$: time at which train t starts the occupation of TDS υ along route r;

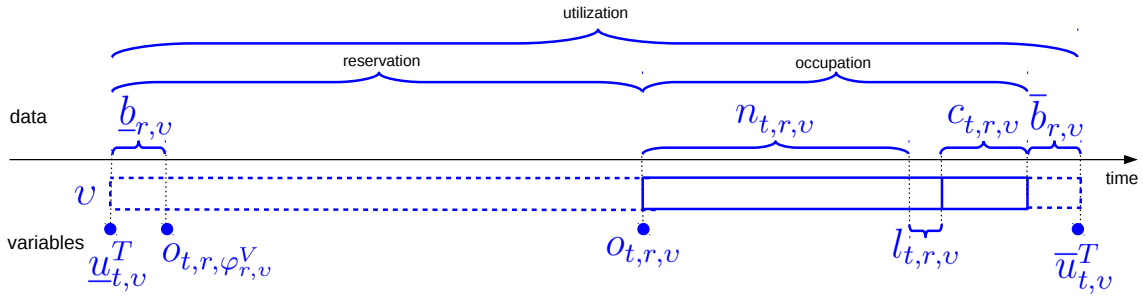


Figure 3.7: Main data and variables concerning the utilization of TDS v belonging to route r of train t when the nominal running and clearing times are used.

$l_{t,r,v}$: longer stay of train t's head on TDS v along route r due to dwell time or a scheduling decision;

for all pairs of $t \in T$ and $v \in \Upsilon^T_t$:

$\underline{u}^T_{t,v}, \overline{u}^T_{t,v}$: time at which TDS v starts and ends being utilized by train t.

for all pairs of $t \in T^O$ and $s \in S_t$:

$\delta^{\underline{d}}_{t,s}, \delta^{\overline{d}}_{t,s}$: early/late departure time difference of train t at control point s, with respect to the train-path request;

$\delta^{\underline{a}}_{t,s}, \delta^{\overline{a}}_{t,s}$: early/late arrival time difference of train t at control point s, with respect to the train-path request.

for all $t \in T^O$:

$\delta^{\underline{\alpha}}_t, \delta^{\overline{\alpha}}_t$: early/late entrance time difference of train t, with respect to the train-path request;

$\delta^{\underline{\Omega}}_t, \delta^{\overline{\Omega}}_t$: early/late exit time difference of train t, with respect to the train-path request.

Binary variables:

for all pairs of $t \in T$ and $r \in R_t$:

$$x_{t,r} = \begin{cases} 1 & \text{if train } t \text{ uses route } r, \\ 0 & \text{otherwise.} \end{cases}$$

for all triplets of $t, t' \in T$ such as $\iota_t < \iota_{t'}$ and $\upsilon \in \bigcup_{r \in R_t} \Upsilon_r^R \cap \bigcup_{r \in R_{t'}} \Upsilon_r^R$:

$$y_{t,t',\upsilon} = \begin{cases} 1 & \text{if train } t \text{ utilizes } \upsilon \text{ before train } t' \ (t \prec t'), \\ 0 & \text{otherwise.} \end{cases}$$

Figure 3.7 displays a graphical representation of the main data and variables concerning the utilization of TDS υ belonging to route r of train t.

3.4.4 Objective Function

As described in Section 3.4, we consider two approaches in RECIFE-TTP for minimizing the time differences with respect to the set of train-path requests: "inner" and "outer".

On the one hand, the "inner" approach consists in maintaining the timetable of the control zone as close as possible to the set of train-path requests. More specifically, the "inner" objective function (3.1), or simply "i", minimizes the weighted sum of departures and arrival time differences at the train's origin, destination and intermediate stops, i.e., control points of a train's itinerary, with respect to the desired schedules defined in the train-path requests. The weights in the sum indicate the priority which may have to be considered for trains and intermediate stops.

$$\min \sum_{t \in T^O, s \in S_t} w_{t,s}^{d} \delta_{t,s}^{d} + w_{t,s}^{\overline{d}} \delta_{t,s}^{\overline{d}} + w_{t,s}^{a} \delta_{t,s}^{a} + w_{t,s}^{\overline{a}} \delta_{t,s}^{\overline{a}}. \tag{3.1}$$

On the other hand, the "outer" approach minimizes the impacts that the timetable of a given control zone may produce in the adjacent ones. Hence, we define the

"outer" objective function (3.2), or simply "o", as the minimization of the weighted sum of entrance and exit time differences of trains with respect to the set of train-path requests. The weights in the sum indicate the impact of the trains may have in adjacent control zones.

$$\min \sum_{t \in T^O} w_t^{\alpha} \delta_t^{\alpha} + w_t^{\overline{\alpha}} \delta_t^{\overline{\alpha}} + w_t^{\Omega} \delta_t^{\Omega} + w_t^{\overline{\Omega}} \delta_t^{\overline{\Omega}}. \tag{3.2}$$

3.4.5 Constraints

Time Related Constraints:

The start time of a TDS v occupation by train t along a route r is zero if the route itself is not used, as shown in Constraints (3.3).

$$o_{t,r,v} \leq M x_{t,r} \qquad \forall t \in T, r \in R_t, v \in \Upsilon_r^R. \tag{3.3}$$

Constraints (3.4) establish that a train t starts occupying a TDS v along a route r after spending in its preceding TDS $\varphi_{r,v}^P$ the corresponding running time, provided that $\varphi_{r,v}^P$ does not belong to a control point $s \in S_t$. Otherwise, the occupation time of v starts after t spends in $\varphi_{r,v}^P$ its corresponding running time plus its longer stay $l_{t,r,v}$, as defined in Constraints (3.5).

$$\begin{aligned} o_{t,r,v} &= o_{t,r,\varphi_{r,v}^P} + n_{t,r,\varphi_{r,v}^P} x_{t,r} \\ \forall t \in T, r \in R_t, v \in \Upsilon_r^R &: \varphi_{r,v}^P \notin \bigcup_{s \in S_t} \Upsilon_{t,s}^S. \end{aligned} \tag{3.4}$$

$$\begin{aligned} o_{t,r,v} &= o_{t,r,\varphi_{r,v}^P} + l_{t,r,\varphi_{r,v}^P} + n_{t,r,\varphi_{r,v}^P} x_{t,r} \\ \forall t \in T, r \in R_t, v \in \Upsilon_r^R &: \varphi_{r,v}^P \in \bigcup_{s \in S_t} \Upsilon_{t,s}^S. \end{aligned} \tag{3.5}$$

The longer stay of a train t along route r in a TDS v belonging to a control point $s \in S_t$ depends on the scheduling mode. On the one hand, for FixDwell, the dwell times at intermediate stations in the train-path requests can not be modified. Therefore, we define Constraints (3.6) to ensure that the longer stay $l_{t,r,v}$ must be equal to the difference between the departure and arrival times at control point s defined in the train-path requests. On the other hand, in VarDwell, the dwell time can be modified as long as the minimum dwell time $p_{t,s}$ is respected. Thus, Constraints (3.7) ensure that the longer stay $l_{t,r,v}$ must be at least equal to $p_{t,s}$.

$$l_{t,r,v} = (d_{t,s} - a_{t,s})x_{t,r}$$
$$\forall t \in T^O, s \in S_t, r \in R_t, v \in \Upsilon^S_{t,s} \cap \Upsilon^R_r \setminus \{\varphi^\Omega_r\}. \tag{3.6}$$

$$l_{t,r,v} \geq p_{t,s}x_{t,r}$$
$$\forall t \in T^O, s \in S_t, r \in R_t, v \in \Upsilon^S_{t,s} \cap \Upsilon^R_r \setminus \{\varphi^\Omega_r\}. \tag{3.7}$$

For two OTs t and t' using the same rolling stock, with $t \in T^B_{t'}$ ($t' \prec t$), at least a separation time $\widehat{p}_{t',t}$ must exist between the arrival at destination of t' and the departure from origin of t, as defined in Constraints (3.8).

$$\sum_{r \in R_t} o_{t,r,\varphi^\alpha_r} \geq \sum_{r' \in R_{t'}} (o_{t',r',\varphi^\Omega_{r'}} + n_{t',r',\varphi^\Omega_{r'}} x_{t',r'} + \widehat{p}_{t',t} x_{t',r'})$$
$$\forall t, t' \in T^O, t \in T^B_{t'}. \tag{3.8}$$

On the one hand, the arrival/departure time differences used in the "i" objective function are defined in Constraints (3.9) to (3.12). The arrival time difference with respect to the desired one of train $t \in T^O$ arriving to control point $s \in S_t$ is defined as the highest non-negative value between an early and a late arrival of t at s, i.e., the desired minus the scheduled arrival times (3.9), or the scheduled minus the desired arrival times (3.10). Note that t is considered as arrived to s only after it has spent

in the TDS $v \in \Upsilon^S_{t,s}$ the corresponding running time, i.e., it stops at the end of the v. Analogously, Constraints (3.11) and (3.12) are defined to determine the departure time difference with respect to the desired schedule of t departing from s.

$$\delta^{\underline{a}}_{t,s} \geq a_{t,s} - \sum_{\substack{r \in R_t: \\ v \in \Upsilon^R_r \cap \Upsilon^S_{t,s}}} (o_{t,r,v} + n_{t,r,v} x_{t,r}) \qquad \forall t \in T^O, s \in S_t. \tag{3.9}$$

$$\delta^{\overline{a}}_{t,s} \geq \sum_{\substack{r \in R_t: \\ v \in \Upsilon^R_r \cap \Upsilon^S_{t,s}}} (o_{t,r,v} + n_{t,r,v} x_{t,r}) - a_{t,s} \qquad \forall t \in T^O, s \in S_t. \tag{3.10}$$

$$\delta^{\underline{d}}_{t,s} \geq d_{t,s} - \sum_{\substack{r \in R_t: \\ v \in \Upsilon^R_r \cap \Upsilon^S_{t,s}}} o_{t,r,\varphi^N_{r,v}} \qquad \forall t \in T^O, s \in S_t. \tag{3.11}$$

$$\delta^{\overline{d}}_{t,s} \geq \sum_{\substack{r \in R_t: \\ v \in \Upsilon^R_r \cap \Upsilon^S_{t,s}}} o_{t,r,\varphi^N_{r,v}} - d_{t,s} \qquad \forall t \in T^O, s \in S_t. \tag{3.12}$$

On the other hand, the entrance/exit time differences used in the "o" objective function are defined in Constraints (3.13) to (3.16). The entrance time difference of train $t \in T^O$ is defined as the highest non-negative value between and early and a late entrance of t to the control zone, as is defined in Constraints (3.13) and (3.14), respectively. The same principle is applied to determine the exit time difference of t leaving the control zone, leading to the definition of Constraints (3.15) and (3.16).

$$\delta^{\underline{\alpha}}_t \geq \underline{e}_t - \sum_{r \in R_t} o_{t,r,\varphi^\alpha_r} \qquad \forall t \in T^O. \tag{3.13}$$

$$\delta^{\overline{\alpha}}_t \geq \sum_{r \in R_t} o_{t,r,\varphi^\alpha_r} - \underline{e}_t \qquad \forall t \in T^O. \tag{3.14}$$

$$\delta^{\underline{\Omega}}_t \geq \overline{e}_t - \sum_{r \in R_t} (o_{t,r,\varphi^\Omega_r} + n_{t,r,\varphi^\Omega_r} x_{t,r}) \qquad \forall t \in T^O. \tag{3.15}$$

$$\delta_t^{\overline{\Omega}} \geq \sum_{r \in R_t} (o_{t,r,\varphi_r^{\Omega}} + n_{t,r,\varphi_r^{\Omega}}) - \overline{e}_t \qquad \forall t \in T^O. \tag{3.16}$$

Routing Related Constraints:

Constraints (3.17) establish that exactly one route must be used for each train. The location coherence between two trains using the same rolling stock is ensured in Constraints (3.18). More precisely, two trains t and t', where $t \in T_{t'}^B$ ($t' \prec t$), must use a route $r \in R_t$ and $r' \in R_{t'}$, respectively, such that the last TDS of r' ($\varphi_{r'}^{\Omega}$) corresponds to the first TDS of r (φ_r^{α}).

$$\sum_{r \in R_t} x_{t,r} = 1 \qquad t \in T. \tag{3.17}$$

$$\sum_{\substack{r' \in R_{t'}: \\ \varphi_{r'}^{\Omega} = \varphi_r^{\alpha}}} x_{t',r'} = \sum_{\substack{r \in R_t: \\ \varphi_r^{\alpha} = \varphi_{r'}^{\Omega}}} x_{t,r} \qquad \forall t, t' \in T^O, t \in T_{t'}^B. \tag{3.18}$$

Capacity Related Constraints:

A train t's utilization of a TDS υ starts as soon as t starts occupying its reference TDS $\varphi_{r,\upsilon}^V$ along the selected route r, minus the formation time, provided that t does not use the same rolling stock of another train, or υ does not belong to the first block section of r, as defined in Constraints (3.19). Otherwise, the utilization of υ starts before that moment. Indeed, the concerned TDS must remain utilized while a turnaround, split or join operation takes place. To do so, Constraints (3.20) impose that the utilization of the first block's TDSs of a train t resulting from t' ($t' \prec t$), starts before the occupation of t's first TDS.

Moreover, Constraints (3.21) establish that the utilization of υ lasts until t finishes utilizing it along any $r \in R_t$ plus the release time. The utilization time $\ddot{u}_{t,r,\upsilon}^T$, defined in Constraints (3.22), includes: First, the running time and longer stays of all TDSs between $\varphi_{r,\upsilon}^V$ and υ. Second, the longer stay on all TDSs υ' such that $\upsilon \in \Upsilon_{t,r,\upsilon'}^C$. This is done to take into account the longer stay spent by t into the forthcoming

TDSs in which the presence of t's head implies that t also occupies υ due to its length. Third, the adequate clearing time of t over υ.

$$\underline{u}^T_{t,\upsilon} = \sum_{\substack{r \in R_t: \\ \upsilon \in \Upsilon^R_r}} (o_{t,r,\varphi^V_{r,\upsilon}} - \underline{b}_{r,\upsilon} x_{t,r})$$

$$\forall t \in T, \upsilon \in \bigcup_{r \in R_t} \Upsilon^R_r : (\nexists t' \in T : t' \in T^B_t) \vee (\varphi^V_{r,\upsilon} \neq \varphi^\alpha_r). \tag{3.19}$$

$$\underline{u}^T_{t,\upsilon} \leq \sum_{\substack{r \in R_t: \\ \upsilon \in \Upsilon^R_r}} (o_{t,r,\varphi^V_{r,\upsilon}} - \underline{b}_{r,\upsilon} x_{t,r})$$

$$\forall t \in T^O, \upsilon \in \bigcup_{r \in R_t} \Upsilon^R_r : (\exists t' \in T : t \in T^B_{t'}) \wedge (\varphi^V_{r,\upsilon} = \varphi^\alpha_r). \tag{3.20}$$

$$\overline{u}^T_{t,\upsilon} = \sum_{\substack{r \in R^t: \\ \upsilon \in \Upsilon^R_r}} (o_{t,r,\varphi^V_{r,\upsilon}} + \ddot{u}^T_{t,r,\upsilon} + \overline{b}_{r,\upsilon} x_{t,r})$$

$$\forall t \in T, \upsilon \in \bigcup_{r \in R_t} \Upsilon^R_r. \tag{3.21}$$

$$\ddot{u}^T_{t,r,\upsilon} \geq \sum_{\upsilon' \in \ddot{\Upsilon}^R_{r,\varphi^V_{r,\upsilon},\upsilon}} (n_{t,r,\upsilon'} x_{t,r} + l_{t,r,\upsilon'}) + \sum_{\substack{\upsilon' \in \Upsilon^R_r: \\ \upsilon \in \Upsilon^C_{t,r,\upsilon'}}} l_{t,r,\upsilon'} + c_{t,r,\upsilon} x_{t,r}$$

$$\forall t \in T, r \in R^T_t, \upsilon \in \Upsilon^R_r. \tag{3.22}$$

The utilization of a TDS υ by two trains t and t' must not overlap in the following two cases: First, if t and t' do not use the same rolling stock. Second, if t and t' use the same rolling stock and υ does not belong to an extreme (first or last) block section of any route $r \in R_t$ and $r' \in R_{t'}$, which is defined in Constraints (3.23) and (3.24).

The utilization of υ may overlap if two trains use the same rolling stock provided that υ belongs to an extreme block section of any route available for these trains. As explained before, a TDS υ concerned with a turnaround, join or split operation

between trains must remain utilized during this time. Therefore, Constraints (3.25) establish that the utilization of υ by train t, resulting from another train t', i.e., $t \in T^B_{t'}$ ($t' \prec t$), must start no later than the end of υ's utilization by t'. This relation is established as an inequality because more than one train may be concerned. For example, consider a join operation in TDS υ, where trains t' and t'' form train t ($t \in T^B_{t'} \cap T^B_{t''}$) and t' arrives before than t'', in this case, the utilization of υ by t starts no later than υ's end of utilization by t', which is before the end of υ's utilization by t''.

$$\overline{u}^T_{t,\upsilon} - \underline{u}^T_{t',\upsilon} \leq M(1 - y_{t,t',\upsilon}) \qquad \forall t, t' \in T, \iota_t < \iota_{t'}, \upsilon \in \bigcup_{r \in R_t} \Upsilon^R_r \cap \bigcup_{r' \in R_{t'}} \Upsilon^R_{r'} :$$
$$(t \notin T^B_{t'} \wedge t' \notin T^B_t) \vee (\upsilon \notin \bigcup_{r \in R_t} \Upsilon^\alpha_r \cup \Upsilon^\Omega_r \wedge \upsilon \notin \bigcup_{r' \in R_{t'}} \Upsilon^\alpha_{r'} \cup \Upsilon^\Omega_{r'}). \tag{3.23}$$

$$\overline{u}^T_{t',\upsilon} - \underline{u}^T_{t,\upsilon} \leq M y_{t,t',\upsilon} \qquad \forall t, t' \in T, \iota_t < \iota_{t'}, \upsilon \in \bigcup_{r \in R_t} \Upsilon^R_r \cap \bigcup_{r' \in R_{t'}} \Upsilon^R_{r'} :$$
$$(t \notin T^B_{t'} \wedge t' \notin T^B_t) \vee (\upsilon \notin \bigcup_{r \in R_t} \Upsilon^\alpha_r \cup \Upsilon^\Omega_r \wedge \upsilon \notin \bigcup_{r' \in R_{t'}} \Upsilon^\alpha_{r'} \cup \Upsilon^\Omega_{r'}). \tag{3.24}$$

$$\sum_{\upsilon \in \bigcup\limits_{r \in R_t} \{\varphi^\alpha_r\}} \underline{u}^T_{t,\upsilon} \leq \sum_{\upsilon \in \bigcup\limits_{r' \in R_{t'}} \{\varphi^\Omega_{r'}\}} \overline{u}^T_{t',\upsilon} \qquad \forall t, t' \in T : t \in T^B_{t'}. \tag{3.25}$$

3.4.6 Reduction of the number of binary variables

To reduce the number of binary variables y, we apply a boosting technique based on the one proposed by Pellegrini et al. (2015), proven to be effective in terms or computational time reduction. In particular, this technique exploits the fact that the topology of a railway infrastructure frequently imposes that precedence relations between trains must be identical on a set of consecutive TDSs.

Consider as an example two trains t and t' following each other on a single track segment containing 10 TDSs (v_0 to v_9) without any overtaking area. According to our formulation, for a couple of trains t and t' circulating over this track section, 10 y-variables are defined, one for each TDS (y_{t,t',v_0} to y_{t,t',v_9}). However, if t enters before t' in the first TDS, i.e., $y_{t,t',v_0} = 1$, given that the precedence between t and t' can not change along the whole single track segment, the value of the remaining y-variables will be 1 as well. Therefore, instead of defining 10 y-variables, we can use only one $y_{t,t',v_{0-9}}$.

By implementing a pre-processing operation to identify all the locations where this may be applicable, which heavily depends on the infrastructure topology, we may significantly reduce the number of binary variables of RECIFE-TTP. Note that when several routes are available for two trains, the set of TDSs which can be considered through a single y variable must be such under any route choice. In our experiments, this boosting technique reduces the number of binary variables in one instance of 88% in the median, i.e., from 73 K binary variables in the original formulation we obtain 9 K binary variables after the pre-processing operation.

3.5 Solution Approach

To solve REFICFE-TTP we present two algorithms:

- *RECIFE-TTP: Single Route (RT-S)* : This algorithm consists in solving a constrained version of RECIFE-TTP by using a MILP solver. This constrained version considers only one route for each train, i.e., the requested one. This implies that x variables are not defined, and hence, the corresponding constraints are updated accordingly.

- *RECIFE-TTP: All Routes (RT-A)* : This algorithm consists in solving the complete formulation described in Section 3.4 by using a MILP solver.

Note that both algorithms support the two scheduling modes of RECIFE-TTP: FixDwell and VarDwell. To differentiate them we add the suffix "f" and "v" for the FixDwell and VarDwell formulations, respectively. Moreover, we use the "inner" and "outer" objective functions for each mode of the algorithms. To identify which objective function is used, we simply add "i" or "o" as a second suffix to the abbreviated names. Therefore, in total, four variants are considered for each algorithm: RT-S-fi, RT-S-fo, RT-S-vi, RT-S-vo, RT-A-fi, RT-A-fo, RT-A-vi and RT-A-vo. Furthermore, we use an "x" to denote both possible variants of an algorithm, e.g., RT-A-fx refers to both RT-A-fi and RT-A-fo, RT-X-fi refers to both RT-S-fi and RT-A-fi and RT-X-xx refers to all 8 algorithm variants.

For each algorithm variant, henceforward simply called algorithms, we establish a *whole time limit*. This is the total wall clock time available for a single execution of the algorithm. After this time, the best feasible solution (if any) is returned. Note that a solution is considered optimal only if the algorithm completes its optimality proof.

3.6 Experiments & Results

The algorithms were implemented in C++ using the Concert Technology for the IBM ILOG CPLEX v12.6 MILP solver (IBM, 2015) and the RECIFE Class Manager Library (RCM-LIB). We developed RCM-LIB as part of RECIFE decision support tools. This C++ library defines a series of data structures that are used for microscopically represent the three main elements of the railway system: infrastructure, rolling stock and timetable. The library implements the CRUD (Create, Read, Update and Delete) functionalities for the persistent storage, in form of XML files, of the previously cited elements. Moreover, it offers a number of classes and methods to easily manipulate the stored data. It performs as well several pre-processing operations such as the calculations needed to reduce the number of variables of RECIFE-TTP. This library can be used as well for further contributions to RECIFE

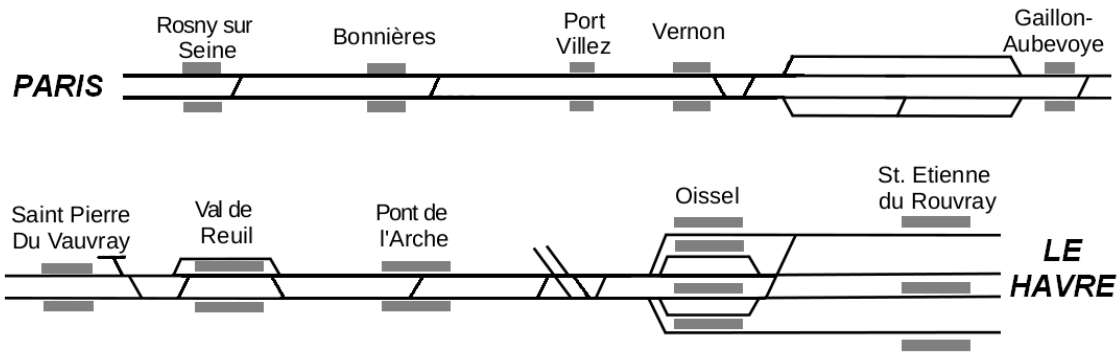


Figure 3.8: Section of the Paris - Le Havre corridor.

decision support tools. More details concerning RCM-LIB can be found in Arenas et al. (2016a).

The experiments were performed on a computer with eight Intel Xeon 3.5 Ghz processors and 128 GB RAM. The whole time limit is set to 4 hours.

3.6.1 Case Study description

We model a section of the Paris - Le Havre corridor which is used by mixed traffic: intercity, regional, high speed and freight trains. The control zone, represented in Figure 3.8, covers a distance of approximately 70 Km, between the stations of Rosny sur Seine and St. Etienne du Rouvray. In this control zone, we identify the following elements: 10 stations, 501 TDSs, 487 block sections, 61 types of rolling stock.

The set of train-path requests was provided by SNCF Réseau, the French IM. These train-path requests actually consisted in the 2012 timetable implemented on the considered control zone, where a microscopic validation had not been performed. The train-path requests concerns 219 trains for one day of operations, 3459 different routes for all trains which represents in mean 15 routes per train (each train has at least 1 route and at most 53). No information is available on the priorities of these trains. Hence, we consider weights equal to 1 for all trains at all control points for both the "i" and "o" objective functions.

3.6.2 Instances

The trains in an instance are determined by the definition of a time *horizon*: they are the trains in the set of train-path requests that are desired to enter the control zone between the start and end times of the horizon.

We generate a set of 50 instances by randomly defining the start time of the horizon with a fixed duration of 4 hours. Each instance is solved by all algorithms, therefore, in total, 400 executions are performed.

The main characteristics of the instances are resumed in Table 3.1. The first part concerns all 50 instances while the second part classes the instances according to their traffic density. To classify the instances we observed the minimum and maximum number of trains considered in all instances, which are 27 and 81, respectively. Then we divide this range into three equal parts which we label "low", "medium" and "high" traffic density, containing up to 45, between 46 and 63, and more than 64 trains, respectively. For each instance class, we report the number of instances, the median number of trains present and the median size of RT-A-vi MILP formulations.

Class	Ins.	Tra.	Cont.Var.	Bin.Var.	Constr.
all	50	68	162 K	9 K	355 K
Traffic Density					
low	7	38	76 K	2 K	141 K
medium	10	59	135 K	6 K	275 K
high	33	76	194 K	12 K	445 K

Table 3.1: Characteristics of the Instances. (Ins.: Number of instances, Tra.: Median number of trains, Cont.Var., Bin.Var. and Constr.: Median number of continuous variables, binary variables and constraints, respectively, in RT-A-vi MILP formulations).

3.6.3 Results

The results of the experiments are reported in Tables 3.2, 3.3 and 3.4. We use, for all tables, the same instance classifications as described for Table 3.1: all instances and instances classed by their traffic density. Each table, contains two sub-tables, one for each group of algorithms using the same objective function, i.e., "inner" and "outer".

We focus first on Table 3.2, where we report for each instance class: % Feasible Sol: the percentage of instances where a feasible solution is obtained by the algorithms. Wall Time: the mean wall clock time, in seconds, required by the algorithms before stopping, either because they found a solution and proved its optimality or because the whole time limit elapsed.

We remark in the first performance indicator of Tables 3.2a and 3.2b that a feasible solution is found for a high percentage of instances by all algorithms within the fixed whole time limit. This high percentage of feasible solutions demonstrates that a microscopic approach can be effectively used in portions of the infrastructure larger than specific nodes. Furthermore, if we compare the percentage of feasible solutions obtained for the instances classed by traffic density, we observe that all algorithms obtain a feasible solution for 100% of the instances containing less than 63 trains. On high density instances, i.e., with 64 trains or more, only RT-A-fi and RT-A-fo are unable to find a feasible solution for 21% and 12% of these instances, respectively.

Concerning the mean wall clock execution time of the algorithms, reported as the second performance indicator in Tables 3.2a and 3.2b, we observe a substantial difference between the single route algorithms (RT-S-xx) with respect to the all routes algorithms (RT-A-xx). Indeed, because of the additional routes to consider, the size of the formulation is greatly increased, along with the search space. Note that RT-A-xx algorithms are able to solve low traffic density instances in a relatively short time, 58 seconds in mean. The processing time augments drastically when they deal with

Class	% Feasible Sol				Wall Time (s)			
	RT-S-fi	RT-S-vi	RT-A-fi	RT-A-vi	RT-S-fi	RT-S-vi	RT-A-fi	RT-A-vi
all	100	100	86	100	11	28	12062	9971
Traffic Density								
low	100	100	100	100	0	1	100	25
medium	100	100	100	100	3	3	12998	3770
high	100	100	79	100	15	41	14315	13960

(a) Algorithms with "i" objective function (3.1)

Class	% Feasible Sol				Wall Time (s)			
	RT-S-fo	RT-S-vo	RT-A-fo	RT-A-vo	RT-S-fo	RT-S-vo	RT-A-fo	RT-A-vo
all	100	100	92	100	10	13	12119	10263
Traffic Density								
low	100	100	100	100	0	1	69	36
medium	100	100	100	100	3	6	13021	3770
high	100	100	88	100	15	18	14400	14400

(b) Algorithms with "o" objective function (3.2)

Table 3.2: Ability of the algorithms to find a feasible solution of an instance within the fixed whole time limit (% Feasible Sol: the percentage of instances where a feasible solution is obtained by the algorithms. Wall Time: the mean wall clock processing time of the algorithms)

medium to high density traffic instances. We can observe as well that, for medium density instances, an important difference of mean processing time exists among the two types of RT-A algorithms, independently of the used objective function: RT-A-fx are considerable slower than RT-A-vx algorithms. This observation, coupled with the fact that RT-A-fx algorithms are the only ones that are unable to find a feasible solution within the time limit, place these algorithms as the least performing ones.

Focusing now on Table 3.3, for each instance class we report: % Optimal Sol: the percentage of instances where the optimal solution is found and proven by the algorithms. GAP: The mean *GAP* of the solutions obtained only for those instances

where a feasible solution is obtained by all algorithms. The GAP indicates the percentage difference between the best feasible solution found by the algorithm within a run and the best bound identified. The higher the GAP, the further the solution may be from the optimum. We consider the GAP as an indicator of the difficulty of an instance for the algorithm: if the GAP is equal to 0 (the optimality is proven), we consider the instance easier than if a positive GAP is returned. The higher the GAP, the more difficult the instance. Recall that the optimal solution of RT-S-xx algorithms may be different from the one of RT-A-xx ones, since only one route is considered for each train.

Class	% Optimal Sol				GAP			
	RT-S-fi	RT-S-vi	RT-A-fi	RT-A-vi	RT-S-fi	RT-S-vi	RT-A-fi	RT-A-vi
all	100	100	16	38	0	0	34	31
Traffic Density								
low	100	100	100	100	0	0	0	0
medium	100	100	10	100	0	0	13	0
high	100	100	0	6	0	0	51	51

(a) Algorithms with "i" objective function (3.1)

Class	% Optimal Sol				GAP			
	RT-S-fo	RT-S-vo	RT-A-fo	RT-A-vo	RT-S-fo	RT-S-vo	RT-A-fo	RT-A-vo
all	100	100	16	34	0	0	40	33
Traffic Density								
low	100	100	100	100	0	0	0	0
medium	100	100	10	100	0	0	28	0
high	100	100	0	0	0	0	54	52

(b) Algorithms with "o" objective function (3.2)

Table 3.3: Ability of the algorithms to find and prove the optimal solution of an instance (% Optimal Sol: the percentage of instances where the optimal solution is found and proven by the algorithms. GAP: the mean percentage difference between the best feasible solution found by the algorithm within a run and the best bound identified.)

We observe on Tables 3.3a and 3.3b that RT-S-xx algorithms prove the optimality of their solutions for all instances whereas, in mean, only 16% and 36% of the instances are solved to optimality by RT-A-fx and RT-A-vx algorithms, respectively. Focusing on the instances classed by traffic density, we remark that the optimal solution is always found by RT-A-vx for medium density instances, but only 10% of this class of instances are optimally solved by RT-A-fx algorithms. No optimal solution is found for any of the high density instances by RT-A-xx algorithms.

Moreover, by observing the different GAP of the solutions obtained by the algorithms, we remark that despite the slight difference existing between the mean GAP obtained by RT-A-fi and RT-A-fo for medium density instances, this difference is practically non-existent for high density ones. Indeed, we observe in general that the GAP value increases along with the size of the formulations, which is directly affected by the number of trains present in the instance and the number of routes to consider.

Results concerning the value of the objective functions are shown in Table 3.4. Independently of the objective function used in the algorithm, i.e., "inner" and "outer" for RT-X-xi and RT-X-xo, respectively, we calculate the value of both objective functions for each algorithm. Therefore, for each instance class and algorithm, we report: Inner Obj.Val.: Mean value of the "inner" objective function. Outer Obj.Val.: Mean value of the "outer" objective function. Bold headers indicate the *actual objective value* of the algorithms. This means that the algorithms below the bold headers use the corresponding objective function. Note that only those instances where a feasible solution is obtained by all algorithms are taken into account to calculate the mean objective values shown in this table.

We first remark that a positive actual value of the objective functions is obtained for all instances, regardless of the used algorithm, even if the optimality was proven. This means that the set of train-path requests is in fact microscopically infeasible. Recall that the train-path requests were provided by SNCF Réseau, the French IM, and they represent an actual timetable of the concerned control zone whose micro-

Class	**Inner Obj.Val**				Outer Obj.Val			
	RT-S-fi	RT-S-vi	RT-A-fi	RT-A-vi	RT-S-fi	RT-S-vi	RT-A-fi	RT-A-vi
all	15280	3974	18849	3296	9213	3664	12098	3020
Traffic Density								
low	10691	553	8439	553	4872	495	3912	495
medium	16121	2852	13026	1855	8882	2600	6779	1776
high	16193	5327	23892	4587	10509	4927	16348	4178

(a) Algorithms with "i" objective function (3.1)

Class	Inner Obj.Val				**Outer Obj.Val**			
	RT-S-fo	RT-S-vo	RT-A-fo	RT-A-vo	RT-S-fo	RT-S-vo	RT-A-fo	RT-A-vo
all	19911	14667	21248	14774	9120	3609	9428	3086
Traffic Density								
low	14394	11301	11417	11567	4808	495	3687	495
medium	20742	13924	16900	15206	8526	2459	6562	1660
high	20956	15735	25121	15399	10366	4757	11803	4202

(b) Algorithms with "o" objective function (3.2)

Table 3.4: A comparison of the objective values of the solutions obtained by the algorithms (Inner Obj.Val.: Mean value of the "inner" objective function. Outer Obj.Val.: Mean value of the "outer" objective function.)

scopical feasibility was not checked. This proves the necessity of using microscopic approaches in the timetabling process to ensure the feasibility of the timetables.

In terms of the actual objective value of the solutions obtained by the algorithms (Table 3.4), we remark that, in most cases, those considering all routes (RT-A-xx) show a lower (better) actual objective value than those considering only one route (RT-S-xx). This is valid for all RT-X-vx algorithms and for those RT-X-fx where the optimal solution is found. Indeed, by considering more routes, the scheduling flexibility is higher and a larger solution space is considered, with potentially lower objective function values.

As an example of this, we evoke the case of a freight train passing through several

stations without stopping. Typically, freight trains are planned to avoid using the main tracks or platforms inside stations as it is commonly assumed that in this way they have a lower impact on passenger trains. Therefore, freight trains are planned to use alternative tracks, if available, in the stations. Note that using an alternative track implies that a switch must be traversed by the train, which in general involves a speed reduction, generating a longer travel time for said train. If at the time when the freight train is scheduled to pass through a station and the main tracks of said station are empty, it is not necessary for that train to use an alternative track. Therefore, when multiple routes are considered, a different route which passes through the main tracks may be assigned to the freight train. By using this different route, the travel time of the train is decreased, which in turn, increases the flexibility of the schedule, e.g., the train can have a longer dwell time in a subsequent station and be able to depart from said station on time.

The advantages of considering multiple routes for the trains can be diminished by the larger processing time needed by RT-A-xx algorithms with respect to the RT-S-xx ones, as shown in Table 3.2. Although a short processing time is not crucial during the timetabling process, because it is carried out in the strategic level of planning (Section 2.3.1), it is important to consider additional techniques to reduce the processing time of RT-A-xx algorithms without compromising the quality of their solutions. The investigation of such techniques is out of the scope of this work and will be considered in future research.

Next, we observe in Table 3.4 that algorithms using a variable dwell time (RT-X-vx) find better solutions than those using a fixed dwell time (RT-X-fx), i.e., a considerably lower mean actual objective value. This can be explained by the fact that restraining the structure of the schedule of a train, e.g., by fixing the dwell times, offers less flexibility for scheduling. This implies a reduced solution space, and possibly a higher optimal objective function value. Considering as well that RT-X-fx algorithms have in general a worse performance in terms of processing time, percentage of feasible solutions and GAP with respect to RT-X-vx, as remarked from

Tables 3.2 and 3.3, we conclude that the algorithms with a fixed dwell time have a bad performance in general.

However, there is an important characteristic in the solutions of RT-X-fx algorithms that should not be disregarded, which is that they preserve any buffer contained in the dwell times of the train-path requests. Indeed, as discussed in Section 3.4, regardless of the method used to define the train-path requests, it is most likely that they already include some time buffers, which is the case of the ones we use for the experiments: the desired stop times (desired departure - possible arrival) were often longer than the minimum dwell times. In this sense, timetables produced using RT-X-fx algorithms are in principle, more robust than those produced by RT-X-vx algorithms, provided the existence of the buffers in the train-path requests.

As discussed in the definition of the objective functions in Section 3.4.4, it is important to consider the impact of the solutions in control zones adjacent to the considered one. Recall that RT-X-xi algorithms minimize the "inner" objective function whereas RT-X-xo minimize the "outer" objective function. We now focus on the comparison of the objective values obtained by the algorithms, which are reported in Table 3.4.

On the one hand, the "outer" objective values obtained by RT-X-xi algorithms remain very close to the corresponding values obtained by RT-X-xo algorithms, which means that the best solutions obtained by RT-X-xi are also rather good solutions when the "outer" objective function is considered. On the other hand, the "inner" objective value obtained by RT-X-xo algorithms is, in general, substantially higher than the corresponding value obtained by RT-X-xi. More specifically, remark that both RT-S-vi and RT-S-vo algorithms have in mean the same "outer" objective value for low traffic instances, i.e., 495. Instead, the "inner" objective value obtained by RT-S-vo (11301) is in mean 20 times larger than the one of RT-S-vi algorithms (553).

To understand the cause of this important difference, as an example, we analyse how a passengers train defined in a particular train-path request of a low traffic

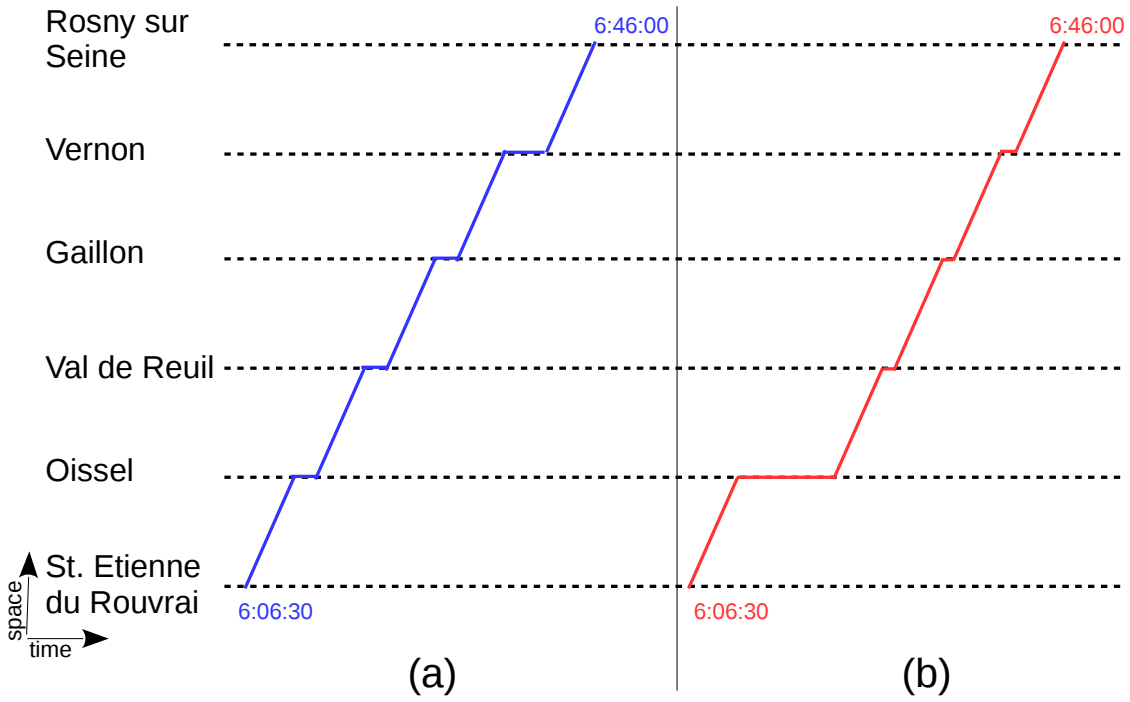


Figure 3.9: Space time diagram example of the scheduling of a train by a RT-S-vo algorithm. (a) the desired schedule of the train defined in its train-path request, (b) the new schedule proposed by the algorithm.

density instance is treated by RT-S-vi and RT-S-vo algorithms. The space time diagram of the referred train is presented in Figure 3.9a. More precisely, the train enters the control zone at "St. Etienne du Rouvrai" and leaves it at "Rosny sur Seine". In the desired schedule of the concerned train-path request, the train is scheduled to enter and exit the control zone at 6:03:30 and 6:46:00, respectively. It has four intermediate stops: "Oissel", "Val de Reuil", "Gaillon" and "Vernon". The dwell times are set to 1 minute for all intermediate stops except for "Vernon", where it lasts 2 minutes. The minimum dwell time is set to 30 seconds for each stop, therefore, the desired dwell times contain time buffers of 30 seconds for the three first stops and 90 for the last one.

In the solution obtained by the RT-S-vo algorithm, the train is scheduled as shown in Figure 3.9b. Note that the train enters and exits the control zone at the exact same times as the desired ones, but the structure of its schedule has changed. Indeed,

the dwell time buffers are eliminated for the last three stations, but substantially increased for the first one. Instead, when the same instance is solved by the RT-S-vi algorithm, the schedule of the train remains unchanged, i.e., the same as the one shown in Figure 3.9a.

Concerning the solution of the RT-S-vo algorithm, the larger dwell time buffer in the first intermediate stop produces a considerable time difference on the arrival/departure times of the subsequent stops of the train, which explains the elevated corresponding "inner" objective value. Recall that the "inner" objective value represents the sum of the arrival/departure time differences, in seconds, at each control point of the train with respect to the desired schedule defined in the train-path request. For this train alone, we observe a contribution of 720 seconds to the "inner" objective value. Instead, since this train's schedule is unchanged in the solution of RT-S-vi, there it has a contribution of 0 to the "inner" objective value.

The solution returned by RT-S-vo is mathematically optimal. Recall that its objective is to minimize the entrance end exit time differences with respect to the desired schedule defined in the train-path request, which for this train is the optimal value of 0, i.e., the train enters and exits the control zone exactly at the desired ones. However, an important number of changes to the schedule of the train inside the control zone are introduced, which is undesirable.

Moreover, note that despite the substantial time differences introduced by RT-S-vo in the beginning of the schedule of the train, the train exits the control zone on time. This is possible because the algorithm eliminates all dwell time buffers in the rest of the train's itinerary. Therefore, the new schedule not only introduces substantial changes to the desired schedule, but also, it is likely to be less robust, provided that the minimum dwell times do not contain any buffer. Indeed, in this work we did not define any mechanism for controlling the buffer times.

3.7 Conclusion

In this chapter we presented in detail the TTP and discussed the main contributions in the literature that tackle this problem. Moreover, we introduced RECIFE-TTP, a MILP formulation of the TTP based on a microscopic infrastructure representation. Timetables produced by RECIFE-TTP are feasible at a microscopic level of details and also highly efficient in terms of railway capacity exploitation.

We performed a series of experiments in a real case scenario of the French railway network. We proposed different variants of the algorithm solving RECIFE-TTP, changing objective functions, scheduling and routing settings.

The results of the experiments proved the effectiveness and practical applicability of RECIFE-TTP. Indeed, the best performing algorithms were able to find a feasible timetable for all the treated instances within an established processing time limit. By considering as well the size of the treated instances, we validated our claim that a microscopic approach can be effectively applied in portions of the infrastructure larger than single stations or junctions. Furthermore, we found that the set of train-path requests used in our experiments, supplied by the French IM is in fact microscopically infeasible. These train-path requests actually consisted in the timetable implemented on the considered control zone. This confirms the necessity of incorporating a microscopic approach in the general timetabling process.

Moreover, after a thorough analysis of the results we were able to identify some important characteristics of the algorithms and the produced timetables. We discovered that algorithms where the dwell times are fixed present a lower performance, in terms of processing time and quality of solutions, compared to those where the dwell time can be modified, however, they are more likely to conserve the buffer times which are important for the robustness of a timetable. We also found out that algorithms minimizing the time differences at every control point of the trains with respect to the initial timetable, i.e., using the "inner" objective function, are

also quite effective in reducing the impact that said trains have in adjacent control zones. Furthermore, algorithms considering multiple routes for the trains provide more scheduling flexibility and better solutions at the cost of considerable higher processing times. In overall, the best performing algorithms where those considering multiple routes, allowing the modification of the dwell times and implementing the "inner" objective function.

A first research perspective is the exploration of different techniques that would allow a reduction of the computational time needed to solve larger instances of the problem without compromising the quality of the solutions. Indeed, solving microscopic approaches to optimality may be very costly in terms of computational time for instances containing hundreds of trains in large portions of the railway infrastructure.

Furthermore, some of the proposed algorithms produced timetables that, despite being mathematically feasible, or even optimal, may not be robust in practice, i.e., the schedules of some trains presented large dwell times at the beginning of their schedule and the rest of its itinerary was set to the theoretical minimal value. The structure of a train schedule should be regular and present a series of incremental time buffers to ensure its robustness. Therefore, we deem important that future work include a deeper study on the specification of time buffers. These should be large enough to improve the robustness of the timetable without excessively reducing the efficiency of the capacity exploitation.

Finally, as mentioned in Section 2.6.3, microscopic approaches need to be integrated with macroscopic ones to produce country-size timetables. Indeed, the enlargement of the scope of the microscopic component to areas of the infrastructure larger than single stations may have an impact on the performance of the integrated approach. This impact may be positive in the sense that the number of macro-micro iterations can be reduced, accelerating the convergence to the solution. However, a negative impact may be also expected if the computational time required by the microscopic component becomes too large. Future studies will be necessary to un-

derstand which of these impacts is stronger.

Chapter 4

Timetable rearrangement to cope with Maintenance Activities

4.1 Introduction

As introduced in Section 2.5, MAs are necessary to maintain the good state of the railway infrastructure, allowing trains to circulate safely and fluidly, thus ensuring the availability of the railway capacity. However, while performing MAs, the train circulations in the concerned locations are impacted. This impact depends on the type of MA performed, which in general implies circulation interdiction on some track segments and temporary speed limitations on neighbouring ones. This means that during the performance of MAs the available capacity is reduced. Moreover, most MAs require at least one *Maintenance Train (MT)*. These are particular trains whose rolling stock is specifically equipped to perform maintenance tasks or to transport maintenance materials. The circulations of these MTs may also impact other trains.

Recall from Section 2.4 that timetables are typically elaborated by considering the track unavailability periods due to MAs and they are considered as feasible if all planned train circulations are conflict-free. To guarantee the feasibility of a timetable, the minimum separation between trains is often based on an underesti-

mation of the actual capacity of the infrastructure during the planning process. This often brings to an inefficient capacity exploitation, since more capacity than what strictly necessary is allocated to each train.

Furthermore, unplanned MAs may be introduced during pre-operational resource reallocation activities in the operational level of the general railway service planning process (Section 2.3.1). This may be caused by an accident or tracks malfunction. The inclusion of unplanned MAs may require rearrangements to the timetable. In the practice, these rearrangements are usually made either by hand, based on the experience of the dispatchers, or by resorting to some optimization tool, based on macroscopic aspects of the infrastructure. In both cases, an efficient exploitation of the capacity is likely not to be achieved.

In this chapter, we propose a microscopic formulation that allows the insertion of MAs into an existing timetable while guaranteeing its feasibility over an area that covers several stations linked by railway lines. More specifically, we present RECIFE-MAINT: a MILP formulation to perform rearrangements on planned train circulations while minimizing the scheduled time deviations with respect to the existing timetable. RECIFE-MAINT has been developed as an extension of RECIFE-TTP, which is thoroughly presented in Section 3.4. RECIFE-MAINT takes into account specific aspects related to MAs that are often disregarded in the literature, such as temporary speed limitations and planning of MTs.

Moreover, we present three algorithms founded on RECIFE-MAINT to solve the problem of rearranging a timetable to cope with MAs. We use these algorithms in the experimental analysis to solve instances based on a case study of the French railway network. Furthermore, we compare the performance of these algorithms with an emulation of the current practice. The contents of this chapter are based on our previous research works presented in Arenas et al. (2016b) and Arenas et al. (2016c).

This chapter is organized as follows. Section 4.2 reviews the related scientific literature. Section 4.3 details the problem of timetable rearrangement to cope with

MAs and the current practice to solve this problem. Section 4.4 presents the complete mathematical formulation of RECIFE-MAINT. Section 4.5 describes the algorithms we propose. Section 4.6 starts by describing a real world case study, which is then used to perform experiments, then reports and discusses the results obtained. Finally, Section 4.7 enlists the conclusions and perspectives of this chapter.

4.2 Literature review

Whereas a large number of works in the scientific literature deal with the TTP (Section 3.2), only a few deal with the unavailability periods of track segments caused by MAs. We classify these contributions into three categories according to the manner in which they deal with MAs:

- Fixed Timetables, Variable Maintenances (FTVM),
- Variable Timetables, Fixed Maintenances (VTFM) and
- Variable Timetables, Variable Maintenances (VTVM).

In the first category, FTVM, we group the contributions where MAs are scheduled by considering the train timetable, but without modifying it. This means that the generated maintenance plan can have conflicts with the timetable, and these conflicts are left to be solved later.

In Higgins (1998), the authors propose a mathematical formulation based on a macroscopic representation of the infrastructure. The objective is to determine the best scheduling for MAs and maintenance crews, while minimizing the disruption to the trains in the timetable and the amount of time needed to complete the MAs. They propose a Tabu Search algorithm.

Budai et al. (2004) propose several heuristic approaches to find near optimal track utilization intervals for carrying out preventive MAs while minimizing the inconvenience for the railway undertaking and the infrastructure utilization time.

Their model is based on a macroscopic representation of the infrastructure, and allows to cluster several MAs together into maintenance packages.

Another work in this category is the one developed by Peng et al. (2011). In this work, the authors propose a time-space network model to schedule MAs while minimizing the total travel costs of the maintenance crews as well as the inconvenience for the railway undertaking. They use an iterative heuristic to solve the problem.

The second category, VTFM, includes the contributions in which MAs are fixed and the timetables are created or modified to cope with them.

In Caprara et al. (2006), the authors present a graph-based mathematical formulation to deal with the timetabling problem while considering MAs. In the author's macroscopic representation of the infrastructure, MAs are modelled as periods of time in which the tracks connecting two stations are unavailable. The authors use a Lagrangian heuristic algorithm.

Another example in this category is the paper by Sourd (2010). The author presents SIOUCS, a tool that optimizes the crossing of trains in a line segment with two tracks, where one of the tracks is unavailable due to an incident or MA. This means that train movements are coordinated to use the single track, in the form of batteries. A battery is a group of trains circulating in the same direction. The optimization criterion is the reduction of the total delay experienced by trains. The infrastructure is represented macroscopically, and a headway time is defined for trains forming the same battery. The solution approach is based on dynamic programming and scheduling theory. This tool has been deployed as decision support module for dispatchers of the national control center of the French railway network.

The third category, VTVM, collects the works where both timetables and MAs are scheduled at the same time. This category can be viewed as a generalisation of the previous ones.

In the formulation presented by Albrecht et al. (2013), trains have a minimal departure time and their routes across the infrastructure can be changed. MAs have

a desired start and end dates, they can be divided into subtasks if necessary and they have a lateness cost function. The solution approach is based on a combination of heuristic methods. Basically, it schedules track utilizations (by trains or subtasks belonging to a MA) one at the time. Some random perturbations are added by the algorithm to create a variety of solutions from which one is chosen. The objective function is the minimization of both the total delay of trains and the accumulated costs of MAs.

The authors of Forsgren et al. (2013) present a mathematical formulation that is solved using a MILP solver. In their formulation, trains can be cancelled, re-routed and rescheduled. MAs, instead, cannot be cancelled, but they can be rescheduled within a predefined time window. The infrastructure is represented macroscopically through a graph where stations or junctions are vertexes and tracks are edges. The capacity of the stations is also taken into account. The objective is to generate a feasible timetable, while minimizing the number of cancelled trains and the total accumulated delay.

The problem we tackle in this chapter can be classed in the second category: VTFM. Differently from the reviewed works we use a microscopic representation of the infrastructure. Moreover, we consider specific constraints imposed by MAs, as temporary speed limitations and the schedule of MTs. The microscopic representation and these constraints are never considered in the cited literature for this problem.

4.3 Problem description

As it was discussed in Section 2.3.1, typically the maintenance needs of the infrastructure are defined before the conception of train timetables. After this, timetables are conceived taking into account the temporal unavailability of track segments caused by MAs. This is often done considering a macroscopic representation of the infrastructure. It is possible then, that the minimum separation between trains dur-

ing MAs is not accurate, e.g., due to an overestimation of the capacity during the macroscopic timetabling procedure, as discussed in Section 2.6.3. In this case, a timetable rearrangement may be necessary before operations to guarantee its feasibility. The same necessity arises when unexpected maintenance needs emerge in the pre-operations level and additional MAs need to be included into the existing timetables.

For safety reasons, during the execution of some types of MAs, a *Temporary Speed Limitation (TSL)* is imposed on adjacent tracks, e.g., a speed limit of 40 km/h on a track segment where trains normally run at 160 km/h. By applying a TSL on a track segment, the time needed for a train to circulate over this segment, i.e., its running time, is increased. This may require adjustments to the schedule of the train.

Furthermore, the circulation of MTs necessary to perform MAs should also be taken into account. The insertion of these trains will likely require additional alterations to other trains' circulations. Trains planned in the timetable, referred as OTs, must respect the green wave policy, instead, MTs are not required to do so.

There are mainly two types of adjustments that can be made to the trains in a timetable: rescheduling and re-routing. We do not consider train cancellations although it may be an alternative, since in the pre-operational level this is not always an option. Indeed, train cancellations may require negotiations with the RUs, which are out of the scope of this work.

Re-routing decisions of the same nature can be applied to OTs and MTs. As defined in Section 2.6.2, a route is the succession of TDSs that a train can use to circulate from an origin to a destination. The alternative routes for a train depend on both the infrastructure and rolling stock characteristics, e.g., some types of rolling stock are not allowed to run in some parts of the infrastructure. In a timetable, trains are scheduled along with their *default route*. A re-routing decision is then the assignment of a route different from the default one to a train.

The nature of scheduling decisions is different for OTs and MTs. On the one

hand, for OTs, the departure time cannot be advanced, only delayed. The reason for this is that in a pre-operational phase, the timetable is normally already known by the users. For instance, the advance of a departure may lead passengers to miss their train. Moreover, OTs must respect the green wave policy, therefore additional delay may be only introduced to OTs while they are dwelling at some station. Ideally, the modifications to the schedule of OTs are made while trying to preserve the planned times of the initial timetable. On the other hand, since MTs are not bind to a timetable, they can be scheduled as it is best convenient. However, MTs have to be present at MA locations when they start, and leave when the MAs are finished. Provided that this holds, MTs can be scheduled to stop at any signal along the track, to be overtaken by OTs.

To summarize, the problem of rearranging a timetable to cope with MAs can be formalized as follows.

Given:

- A railway infrastructure.
- An initial timetable.
- A set of MAs.

Find a feasible timetable compatible with all MAs where:

- All OTs from the initial timetable are scheduled according to the green wave policy.
- All the required MTs are scheduled.
- All capacity, safety and TSL constraints are respected.
- The scheduled time deviations with respect to the initial timetable are minimized.

For this problem, we think that a microscopic representation of the infrastructure considering larger areas than the ones directly concerned by the MAs is necessary

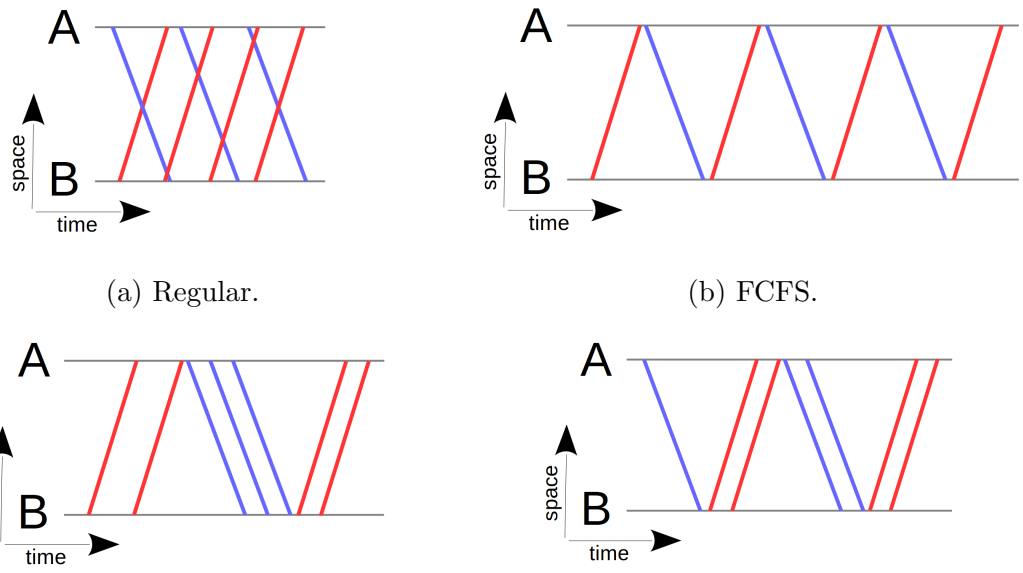
because of the following reasons:

- MAs can occur in small parts of the infrastructure, therefore a way to identify specific track sections is necessary.
- It must be possible to locally re-route trains. Thus, it is essential to be able to distinguish the different routes available for a train and, hence, to exploit all the possibilities offered in the practice by the interlocking system. Even small differences between routes can have a large impact on the quality of the final timetable.
- The MAs limit the railway capacity available for OTs, hence, this capacity should be exploited as efficiently as possible.
- Capacity limitations due to a MA can have consequences on an extended part of the network.
- In this context of scarce capacity, guaranteeing the feasibility of a timetable is particularly important.

4.3.1 Current practice: Train Batteries

The constitution of train batteries is a commonly used approach to deal with the problem presented in the previous section. A *train battery* is a group of trains circulating in the same direction. Scheduling trains in batteries improves the efficiency of the capacity exploitation in a single-track segment with traffic running in both directions. Indeed, train batteries allow the reduction of headway times between trains, compared to the situation in which trains travelling in opposite directions cross in, e.g., a *First-Come First-Served (FCFS)* order.

Figure 4.1 shows an example of train rescheduling. The space-time diagrams represent a small part of a railway infrastructure containing two stations: A and B. These stations are connected by a double-track segment, therefore, in Figure 4.1a trains are scheduled in both directions without any conflict. However, if one track be-



(a) Regular.

(b) FCFS.

(c) Batteries, B-A first.

(d) Batteries, A-B first.

Figure 4.1: Example of train rescheduling using FCFS and batteries approaches. In (a) a space-time diagram shows a regular schedule of seven trains in a double-track segment equipped with an IPCS and connecting stations A and B. Provided that one track becomes unavailable, new schedules are proposed in which trains are rescheduled by applying FCFS (b) and batteries (c and d) approaches.

comes unavailable, e.g., because of an unplanned MA, the trains need to be rescheduled to use the available track without any conflict. This can be done provided that the available track supports opposite direction circulation, i.e., is equipped with an IPCS, as discussed in Section 2.5. Three new train schedules are obtained by applying two approaches: FCFS (Figure 4.1b) and train batteries (Figures 4.1c and 4.1d).

By comparing the different schedules proposed in Figure 4.1, we observe that the train schedules obtained by using train batteries (Figures 4.1c and 4.1d) effectively rescheduled all trains present in Figure 4.1a while utilizing the railway capacity more efficiently than the schedule obtained by FCFS (Figure 4.1b): the same number of trains are scheduled using the infrastructure for a shorter time. Moreover, trains rescheduled in batteries experience a lower delay than the ones rescheduled following a FCFS approach.

By analysing the differences between Figures 4.1c and 4.1d, we can identify three parameters that characterize a schedule obtained by following a batteries approach: The number of train batteries (3 in Figure 4.1c and 4 in Figure 4.1d), the number of trains per battery (2,3,2 and 1,2,2,2 in Figures 4.1c and 4.1d, respectively) and which is the direction of the first battery (B to A versus A to B). Note that in some cases it may be convenient to compose a schedule by considering some batteries of only one train (Figure 4.1d). The settings of these parameters may have a high impact on the quality of the timetable produced.

As an example of the application of the train batteries approach in a real case scenario, we cite the Tours-Bordeaux case study, which is thoroughly described in Brasme (2012).

In 2012, the double-track railway line connecting the cities of Tours and Bordeaux has been subject to a series of heavy renewal MAs. In total, 506 Km of tracks were renewed in this process. The MAs were planned as non-stable, permeable windows of 7 hours (Section 2.5.1) to be performed at nights, between 22h30 and 05h30. During this period of time, approximatively 30 trains per direction were planned to use this part of the infrastructure. The planning of these trains was allowed because the tracks of this line are equipped with IPCSs (Section 2.5). However, their definitive schedules were still to be determined. The problem was, then, to effectively schedule these train circulations while respecting the typical circulation and safety constraints plus the unavailability of the tracks affected by MAs and the TSLs imposed.

Because of the complexity of the problem at hand and the limitations of the available rescheduling tools, the chosen approach was to reschedule the trains using the batteries approach. The number and composition of the train batteries depended on the location of the MA and the initial schedules of the trains. For example, three train batteries, the first one using the single track segment between 10h30 and 00h30 in direction Tours, the second one between 00h30 and 03h30 in direction Bordeaux and the last one between 03h30 and 05h30 in direction Tours.

Although all trains were scheduled using this approach, the capacity exploitation was not very efficient. Indeed, a number of the rescheduled trains were severely delayed with respect to their initial schedule. In some cases this delay attained more than 2h30. After an *a posteriori* analysis of the results, the experts agreed that an optimized microscopic approach would have most likely improved the rearrangement of the timetable.

4.4 Problem formulation

In this section we present RECIFE-MAINT, a MILP formulation to tackle the problem described in Section 4.3. Based on the RECIFE-TTP formulation (Section 3.4), this extension introduces new variables and constraints for the model of MAs, the planning of MTs and the observance of TSLs. In this section, we describe in detail these new elements and the main differences with respect to RECIFE-TTP.

First, as described in Section 3.4, one of the main inputs of RECIFE-TTP is a set of train-path requests, which in turn, may be an interpretation of an actual timetable. In this sense, the input data used by RECIFE-MAINT is equivalent to the one used by RECIFE-TTP. The difference is mostly semantic: the set of train-path requests are interpreted in RECIFE-MAINT as the set of trains actually routed and scheduled in a timetable, the so called initial timetable.

Concerning the modelling of MAs, we denote as *directly affected TDS* the set of consecutive TDSs where the MA takes place. No train circulation is allowed in these TDSs during the realization of the MA. Instead, the *TSL affected TDS* are the ones, typically contiguous to the directly affected TDSs, where train circulations are allowed during the performance of the MA provided that they respect a TSL. A single MA can define different sets of TSL affected TDS. Moreover, a single TDS can be TSL affected by more than one MA. In this case, the lowest speed limitation is imposed. Figure 4.2 shows a representation of the elements described above, note that MA2 imposes a TSL in two sets of TDSs, moreover, tds22 is affected by the

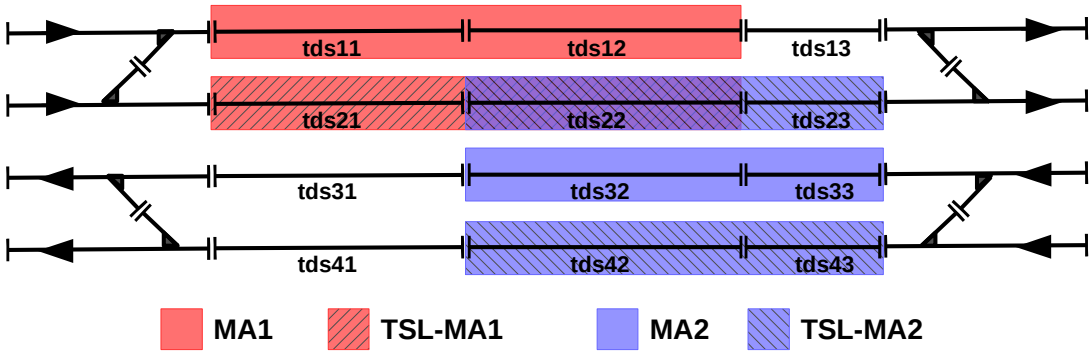


Figure 4.2: Example of two MAs affecting a set of TDSs. MA1 directly affects tds11 and tds12, while imposing a TSL in tds21 and tds22. MA2 directly affects tds32 and tds33, and it imposes a TSL in two set of TDSs: First, tds22 and tds23, and second, tds42 and tds43.

TSL imposed by both MA1 and MA2.

Next, due to the possibility of TSL impositions, at least two different running times exist for each TSL affected TDS utilized by a train. The *nominal running time* and the *restricted running time*, which correspond to the running time of the train traversing the TSL affected TDS in normal conditions and during the execution of a MA, respectively. The same concept applies for the clearing time. Remark that the specific running time to be used, i.e., either nominal or restricted, depends on the time at which the train passes through the TSL affected TDS. This makes the problem intrinsically non-linear but we linearize the concerned constraints with the use of the Big M method.

Finally, different types of constraints are applied to a MT whether it is arriving at or departing from a MA location. On the one hand, the MT must arrive to the MA location at the exact time as the MA begins, i.e., the final arrival time is constrained. On the other hand, the MT must depart from the MA location as soon as the MA is finished, i.e., the departure time is constrained. In the following, we will refer to these two MTs as *inbound MT* and *outbound MT*, respectively.

4.4.1 Assumptions for the MILP notation

The notation used in this section follows the same principles defined in Section 3.4.1. The new symbol to consider is Λ, which refers to MAs.

4.4.2 Data

The input data defined in Section 3.4.2 are also used in RECIFE-MAINT. Some modifications are performed and described along with the additions.

Modifications:

$\underline{e}_t$ and $\overline{e}_t$ are redefined as:

$\underline{e}_t, \overline{e}_t \equiv$ earliest time at which train $t \in T^O$ can be operated, and scheduled arrival to its destination given $\underline{e}_t$, the default route and intermediate stops;

$w^{\underline{d}}_{t,s}$, $w^{\overline{d}}_{t,s}$, $w^{\underline{a}}_{t,s}$, $w^{\overline{a}}_{t,s}$, $w^{\underline{\alpha}}_{t}$, $w^{\overline{\alpha}}_{t}$, $w^{\underline{\Omega}}_{t}$ and $w^{\overline{\Omega}}_{t}$ are replaced by:

$w_{t,s}, w_{t,s} \equiv$ weight associated to one time unit of arrival delay of train $t \in T^O$ at control point $s \in S_t$

$n_{t,r,\upsilon}$ and $c_{t,r,\upsilon}$ are replaced by:

$\underline{n}_{t,r,\upsilon}, \overline{n}_{t,r,\upsilon} \equiv$ nominal and restricted running time of train $t \in T$ along route $r \in R_t$ on TDS $\upsilon \in \Upsilon^R_r$ ($\underline{n}_{t,r,\upsilon}$ and $\overline{n}_{t,r,\upsilon}$ replace $n_{t,r,\upsilon}$ in RECIFE-TTP);

$\underline{c}_{t,r,\upsilon}, \overline{c}_{t,r,\upsilon} \equiv$ nominal and restricted clearing time of train $t \in T$ along route $r \in R_t$ on TDS $\upsilon \in \Upsilon^R_r$ ($\underline{c}_{t,r,\upsilon}$ and $\overline{c}_{t,r,\upsilon}$ replace $c_{t,r,\upsilon}$ in RECIFE-TTP);

Additions:

$\Lambda \equiv$ set of MAs;

$\underline{u}^{\Lambda}_{\lambda}, \overline{u}^{\Lambda}_{\lambda} \equiv$ start and end time of MA $\lambda \in \Lambda$;

$\Upsilon^{\Lambda}_{\lambda} \equiv$ set of directly affected TDS by MA $\lambda \in \Lambda$;

$\widehat{\Upsilon}^{\Lambda}_{\lambda} \equiv$ set of TSL affected TDS during MA $\lambda \in \Lambda$;

$T^{\underline{\Lambda}}_{\lambda}, T^{\overline{\Lambda}}_{\lambda} \equiv$ set of inbound and outbound MTs required for MA $\lambda \in \Lambda$;

4.4.3 Variables

The variables defined in Section 3.4.3 are also used in RECIFE-MAINT. Some modifications are performed and described along with the additions.

Modifications:

$\delta^{\underline{d}}_{t,s}$, $\delta^{\overline{d}}_{t,s}$, $\delta^{\underline{a}}_{t,s}$, $\delta^{\overline{a}}_{t,s}$, $\delta^{\underline{\alpha}}_{t}$, $\delta^{\overline{\alpha}}_{t}$, $\delta^{\underline{\Omega}}_{t}$ and $\delta^{\overline{\Omega}}_{t}$ are replaced by the non-negative continuous variable $\delta_{t,s}$:

for all pairs of $t \in T^O$ and $s \in S_t$:

$\delta_{t,s}$: delay suffered by train t arriving at control point s;

Additions:

Non-negative continuous variables:

for all triplets of $t \in T$, $r \in R_t$ and $\upsilon \in \Upsilon^R_r$:

$\eta_{t,r,\upsilon}$: *actual running time* of train t along route r on TDS υ taking into account the possibly imposed TSL and longer stay;

Binary variables:

for all triplets of $t \in T$, $\lambda \in \Lambda$ and $\upsilon \in \bigcup_{r \in R_t} \Upsilon^R_r \cap \Upsilon^{\Lambda}_{\lambda}$:

$$z_{t,\lambda,\upsilon} = \begin{cases} 1 & \text{if train } t \text{ starts utilizing TDS } \upsilon \text{ before the start of MA } \lambda \ (t \prec \lambda), \\ 0 & \text{otherwise.} \end{cases}$$

To cope with the TSL imposed by a MA λ, three sets of binary variables are defined for all $\upsilon \in \widehat{\Upsilon}^{\Lambda}_{\lambda}$. First, variable f is defined to establish if a train $t \in T$ uses υ before the start of λ, and hence before the start of the speed limitation. Then, variable g is defined to identify the case in which t uses υ while λ is being performed and the speed limitation is active. Finally, variable m is set to determine if t runs through υ

during a TSL imposition, i.e., if at least one g variable is set to one. This variable is called *running time mode*. Recall that υ may be affected by TSLs imposed by several MAs, as shown in the example in Figure 4.2. If a TSL imposition exists, the restricted running time must be used to calculate the actual running time ($\eta_{t,r,\upsilon}$), otherwise, the nominal running time is used. Analogously, either the restricted or the nominal clearing time is used during the calculation of the utilization time.

For all triplets of $t \in T$, $\lambda \in \Lambda$ and $\upsilon \in \bigcup_{r \in R_t} \Upsilon_r^R \cap \widehat{\Upsilon}_\lambda^\Lambda$:

$$f_{t,\lambda,\upsilon} = \begin{cases} 1 & \text{if train } t \text{ starts utilizing TDS } \upsilon \text{ before the start of MA } \lambda \ (t \prec \lambda), \\ 0 & \text{otherwise,} \end{cases}$$

$$g_{t,\lambda,\upsilon} = \begin{cases} 1 & \text{if train } t \text{ uses TDS } \upsilon \text{ during the realization of MA } \lambda, \\ 0 & \text{otherwise,} \end{cases}$$

for all pairs of $t \in T$ and $\upsilon \in \bigcup_{r \in R_t} \Upsilon_r^R$:

$$m_{t,\upsilon} = \begin{cases} 1 & \text{if train } t \text{ runs through } \upsilon \text{ during the imposition of a TSL,} \\ 0 & \text{otherwise.} \end{cases}$$

4.4.4 Objective Function

The objective function minimizes the weighted sum of arrival delays at the OT's destination and intermediate stops, with respect to the initial timetable (4.1).

$$\min \sum_{t \in T^O, s \in S_t} \delta_{t,s} w_{t,s}. \tag{4.1}$$

4.4.5 Constraints

A number of constraints defined for RECIFE-TTP, and described in Section 3.4.5, are also applied in RECIFE-MAINT. Some of these constraints remain unchanged,

others are adapted. In this section we rapidly review the constraints that are reused, then we describe in detail the adapted constraints along with the new ones. Table 4.1 summarizes the constraints differences between RECIFE-TTP and RECIFE-MAINT.

Constraints	**RECIFE-TTP**	**RECIFE-MAINT**
Timing	(3.3) - (3.4) (3.5) (3.6) (3.7) (3.8) (3.9) (3.10) (3.11) (3.12) (3.13) (3.14) (3.15) (3.16) -	(3.3) (4.2) (4.4) (4.5) (4.6) (4.7) (4.8) (4.9) (4.10) (4.11) - (3.7) (4.12) - (4.13) (4.14)
Routing	(3.17) (3.18)	(3.17) (3.18)
Capacity	(3.19) (3.20) (3.21) (3.22) (3.23) (3.24) (3.25)	(3.19) (3.20) (3.21) (4.15) (4.16) (3.23) (3.24) (3.25)
Maintenance	-	(3.17) (4.18) (4.19) (4.20)
TSL	-	(4.21) (4.22) (4.23) (4.24) (4.25) (4.26) (4.27) (4.28)

Table 4.1: Summary of the main differences between the constraints of RECIFE-TTP & RECIFE-MAINT.

The reused timing constraints are: Constraints (3.3), which set to zero the occupation time of a TDS which belongs to an unused route of a train, and Constraints (3.7), which regulate the minimum dwell times of an OT.

The reused routing constraints are: Constraints (3.17), which ensure that only one route is used by a train, and Constraints (3.18), which ensure the coherence

between two trains using the same rolling stock.

The reused capacity constraints are: Constraints (3.19), (3.20) and (3.21), which control the start and end of the utilization time of the TDSs utilized by a train. Additionally, Constraints (3.23), (3.24) and (3.25), which prevent the simultaneous utilization of a TDS by two trains unless the concerned trains are planned to perform a rolling stock balance operation in the concerned TDS.

Next, we detail the new constraints along with the adapted ones.

Timing Constraints:

We defined Constraints (4.2) to prevent a train t to be operated before its earliest operable time $\underline{e}_t$.

$$o_{t,r,\upsilon} \geq \underline{e}_t x_{t,r} \qquad \forall t \in T, r \in R_t, \upsilon \in \Upsilon_r^R. \tag{4.2}$$

Constraints (3.4) and (3.5) are adapted to support the two different running times available in RECIFE-MAINT: A train t starts occupying a TDS υ along a route r after spending in the preceding TDS $\varphi_{r,\upsilon}^P$ the corresponding actual running time $\eta_{t,r,\varphi_{r,\upsilon}^P}$, as defined in Constraints (4.3).

$$o_{t,r,\upsilon} = o_{t,r,\varphi_{r,\upsilon}^P} + \eta_{t,r,\varphi_{r,\upsilon}^P} \qquad \forall t \in T, r \in R_t, \upsilon \in \Upsilon_r^R. \tag{4.3}$$

Then, the actual running time of t along r over υ ($\eta_{t,r,\upsilon}$) has a slightly different definition depending if a longer stay $l_{t,r,\upsilon}$ is indeed allowed in υ or not. Recall that OTs must respect the green wave policy, i.e., longer stays can only exist in their intermediate stations, whereas MTs can suffer longer stays in any TDSs.

On the one hand, the actual running time for TDSs where a longer stay is not possible, defined in Constraints (4.4), (4.5), (4.6) and (4.6), is equal to the running time, depending on the running time mode $m_{t,\upsilon}$ to be used. Here, if $m_{t,\upsilon} = 1$, Constraints (4.4) and (4.5) will be active and the restricted running time will be

imposed. Otherwise, Constraints (4.6) and (4.7) will be active and the nominal running time will be set.

$$\eta_{t,r,\upsilon} \geq \overline{n}_{t,r,\upsilon} x_{t,r} - M(1 - m_{t,\upsilon}) \qquad \forall t \in T^O, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \notin \bigcup_{s \in S_t} \Upsilon_{t,s}^S. \tag{4.4}$$

$$\eta_{t,r,\upsilon} \leq \overline{n}_{t,r,\upsilon} x_{t,r} + M(1 - m_{t,\upsilon}) \qquad \forall t \in T^O, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \notin \bigcup_{s \in S_t} \Upsilon_{t,s}^S. \tag{4.5}$$

$$\eta_{t,r,\upsilon} \geq \underline{n}_{t,r,\upsilon} x_{t,r} - M m_{t,\upsilon} \qquad \forall t \in T^O, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \notin \bigcup_{s \in S_t} \Upsilon_{t,s}^S. \tag{4.6}$$

$$\eta_{t,r,\upsilon} \leq \underline{n}_{t,r,\upsilon} x_{t,r} - M m_{t,\upsilon} \qquad \forall t \in T^O, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \notin \bigcup_{s \in S_t} \Upsilon_{t,s}^S. \tag{4.7}$$

On the other hand, the actual running time for TDSs where a longer stay is possible, defined in Constraints (4.8), (4.9), (4.10) and (4.10), is equal to the sum of the longer stay of t on υ and its running time, depending on the running time mode $m_{t,\upsilon}$ to be used. Here, we apply the same principle concerning the running time mode as explained before.

$$\eta_{t,r,\upsilon} \geq l_{t,r,\upsilon} + \overline{n}_{t,r,\upsilon} x_{t,r} - M(1 - m_{t,\upsilon})$$
$$\forall t \in T, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \in \bigcup_{s \in S_t} \Upsilon_{t,s}^S \vee t \notin T^O. \tag{4.8}$$

$$\eta_{t,r,\upsilon} \leq l_{t,r,\upsilon} + \overline{n}_{t,r,\upsilon} x_{t,r} + M(1 - m_{t,\upsilon})$$
$$\forall t \in T, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \in \bigcup_{s \in S_t} \Upsilon_{t,s}^S \vee t \notin T^O. \tag{4.9}$$

$$\eta_{t,r,\upsilon} \geq l_{t,r,\upsilon} + \underline{n}_{t,r,\upsilon} x_{t,r} - M m_{t,\upsilon}$$
$$\forall t \in T, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \in \bigcup_{s \in S_t} \Upsilon_{t,s}^S \vee t \notin T^O. \tag{4.10}$$

$$\eta_{t,r,\upsilon} \leq l_{t,r,\upsilon} + \underline{n}_{t,r,\upsilon} x_{t,r} - M m_{t,\upsilon}$$
$$\forall t \in T, r \in R_t, \upsilon \in \Upsilon_r^R : \upsilon \in \bigcup_{s \in S_t} \Upsilon_{t,s}^S \vee t \notin T^O. \tag{4.11}$$

Constraints (4.12) are defined in a very similar manner as Constraints (3.8), the main difference is that the actual running time is used.

$$\sum_{r \in R_t} o_{t,r,\varphi_r^\alpha} \geq \sum_{r' \in R_{t'}} (o_{t',r',\varphi_{r'}^\Omega} + \eta_{t',r',\varphi_{r'}^\Omega} + \widehat{p}_{t',t} x_{t',r}) \qquad \forall t, t' \in T^O, t \in T_{t'}^B. \tag{4.12}$$

Differently from RECIFE-TTP, in this formulation OTs can not be advanced, therefore we define Constraints (4.13) to ensure that an OT t can not depart from a control point $s \in S_t$ before its scheduled departure time $d_{t,s}$, .

$$o_{t,r,\varphi_{r,\upsilon}^N} \geq d_{t,s} x_{t,r} \qquad \forall t \in T^O, s \in S_t, r \in R_t, \upsilon \in \Upsilon_{t,s}^S \cap \Upsilon_r^R \setminus \{\varphi_r^\Omega\}. \tag{4.13}$$

An OT t having a set of control points S_t, suffers a delay $\delta_{t,s}$ at least equal to the difference between the actual and the scheduled arrival times at said control point, as defined in Constraints (4.14).

$$\delta_{t,s} \geq \sum_{\substack{r \in R_t: \\ \upsilon \in \Upsilon_r^R \cap \Upsilon_{t,s}^S}} (o_{t,r,\upsilon} + \eta_{t,r,\upsilon} - l_{t,r,\upsilon}) - a_{t,s} \qquad \forall t \in T^O, s \in S_t. \tag{4.14}$$

Capacity Constraints:

The utilization time $\ddot{u}_{t,r,\upsilon}^T$, initially defined in Constraints (3.22), is redefined in Constraints (4.15) and (4.16) to include the actual running times and the adequate clearing times corresponding to the running time mode of train t over TDS υ.

$$\ddot{u}_{t,r,\upsilon}^T \geq \sum_{\upsilon' \in \ddot{\Upsilon}_{r,\varphi_{r,\upsilon},\upsilon}^R} \eta_{t,r,\upsilon'} \quad + \sum_{\substack{\upsilon' \in \Upsilon_r^R : \upsilon \in \Upsilon_{t,r,\upsilon'}^C \\ \wedge (t \notin T^O \vee \upsilon' \in \bigcup_{s \in S_t} \Upsilon_{t,s}^S)}} l_{t,r,\upsilon'} + \bar{c}_{t,r,\upsilon} x_{t,r} - M(1 - m_{t,\upsilon})$$

$$\forall t \in T, r \in R_t^T, \upsilon \in \Upsilon_r^R. \tag{4.15}$$

$$\ddot{\ddot{u}}^T_{t,r,v} \geq \sum_{v' \in \ddot{\ddot{\Upsilon}}^R_{r,\varphi^V_{r,v},v}} \eta_{t,r,v'} \quad + \sum_{\substack{v' \in \Upsilon^R_r : v \in \Upsilon^C_{t,r,v'} \\ \wedge (t \notin T^O \vee v' \in \bigcup_{s \in S_t} \Upsilon^S_{t,s})}} l_{t,r,v'} + \underline{c}_{t,r,v} x_{t,r} - M m_{t,v}$$
$$\forall t \in T, r \in R^T_t, v \in \Upsilon^R_r. \tag{4.16}$$

Maintenance Constraints:

We define Constraints (4.17) and (4.18) to ensure that an inbound MT $t \in T^{\underline{\Lambda}}_\lambda$ arrives at the location of MA λ when λ itself starts, and an outbound MT $t \in T^{\overline{\Lambda}}_\lambda$ departs from the location of λ when λ itself finishes, respectively.

$$\underline{u}^\Lambda_\lambda = \sum_{r \in R_t} o_{t,r,\varphi^\Omega_r} + \eta_{t,r,\varphi^\Omega_r} \qquad \forall \lambda \in \Lambda, t \in T^{\underline{\Lambda}}_\lambda. \tag{4.17}$$
$$\overline{u}^\Lambda_\lambda = \sum_{r \in R_t} o_{t,r,\varphi^\alpha_r} \qquad \forall \lambda \in \Lambda, t \in T^{\overline{\Lambda}}_\lambda. \tag{4.18}$$

The TDS utilizations by a train and a MA must not overlap if the TDS itself belongs to the directly affected TDS list of the MA, as defined in Constraints (4.19) and (4.20).

$$\overline{u}^T_{t,v} - \underline{u}^\Lambda_\lambda \leq M(1 - z_{t,\lambda,v})$$
$$\forall \lambda \in \Lambda, t \in T \setminus (T^{\underline{\Lambda}}_\lambda \cup T^{\overline{\Lambda}}_\lambda), v \in \Upsilon^\Lambda_\lambda \cap \bigcup_{r \in R_t} \Upsilon^R_r. \tag{4.19}$$

$$\overline{u}^\Lambda_\lambda - \underline{u}^T_{t,v} \leq M z_{t,\lambda,v}$$
$$\forall \lambda \in \Lambda, t \in T \setminus (T^{\underline{\Lambda}}_\lambda \cup T^{\overline{\Lambda}}_\lambda), v \in \Upsilon^\Lambda_\lambda \cap \bigcup_{r \in R_t} \Upsilon^R_r. \tag{4.20}$$

Temporary Speed Limitation Related Constraints:

A train t can circulate at nominal speed over a TDS v belonging to the list of TSL affected TDS by a MA λ, only if v's utilization by t takes place while λ is not

being performed, i.e., either if the complete utilization of υ by t takes place before the start of λ, or, if it takes place after the end of λ.

We define Constraints (4.21) and (4.22) to identify whether t starts the utilization of υ before or after the beginning of λ. Next, Constraints (4.23), (4.24), (4.25) and (4.26) are defined to ensure the proper behaviour of binary variable g. This means that for a train t using a TDS $\upsilon \in \bigcup_{r \in R_t} \Upsilon_r^R \cap \widehat{\Upsilon}_\lambda^\Lambda$, the value of $g_{t,\lambda,\upsilon}$ must be set to 0 only in two cases: First, when the utilization of υ by t starts before the beginning of λ ($f_{t,\lambda,\upsilon} = 1$) and finishes before λ starts. Second, when the utilization of υ by t starts after the beginning of λ ($f_{t,\lambda,\upsilon} = 0$) and finishes after the end of λ. Otherwise, $g_{t,\lambda,\upsilon}$ must be set to 1.

On the one hand, Constraints (4.23) and (4.24) are active when $f_{t,\lambda,\upsilon} = 1$; then, they ensure that the value of $g_{t,\lambda,\upsilon}$ is set to 0 only if $\overline{u}_{t,\upsilon}^T < \underline{u}_\lambda^\Lambda$. On the other hand, Constraints (4.25) and (4.26) are active when $f_{t,\lambda,\upsilon} = 0$; then, they ensure that the value of $g_{t,\lambda,\upsilon}$ is set to 0 only if $\overline{u}_\lambda^\Lambda < \underline{u}_{t,\upsilon}^T$.

Constraints (4.27) and (4.28) ensure that the running time mode of t over υ is set to restricted ($m_{t,\upsilon} = 1$) if t utilizes υ during the performance of one or more MA λ that imposes a TSL on υ.

$$\underline{u}_{t,\upsilon}^T - \underline{u}_\lambda^\Lambda \leq M(1 - f_{t,\lambda,\upsilon}) \qquad \forall t \in T, \lambda \in \Lambda : \upsilon \in \widehat{\Upsilon}_\lambda^\Lambda \cap \bigcup_{r \in R_t} \Upsilon_r^R. \tag{4.21}$$

$$\underline{u}_\lambda^\Lambda - \underline{u}_{t,\upsilon}^T < M f_{t,\lambda,\upsilon} \qquad \forall t \in T, \lambda \in \Lambda : \upsilon \in \widehat{\Upsilon}_\lambda^\Lambda \cap \bigcup_{r \in R_t} \Upsilon_r^R. \tag{4.22}$$

$$\overline{u}_{t,\upsilon}^T - \underline{u}_\lambda^\Lambda \leq M(1 - f_{t,\lambda,\upsilon} + g_{t,\lambda,\upsilon}) \qquad \forall t \in T, \lambda \in \Lambda : \upsilon \in \widehat{\Upsilon}_\lambda^\Lambda \cap \bigcup_{r \in R_t} \Upsilon_r^R. \tag{4.23}$$

$$\underline{u}_\lambda^\Lambda - \overline{u}_{t,\upsilon}^T < M(2 - f_{t,\lambda,\upsilon} - g_{t,\lambda,\upsilon}) \qquad \forall t \in T, \lambda \in \Lambda : \upsilon \in \widehat{\Upsilon}_\lambda^\Lambda \cap \bigcup_{r \in R_t} \Upsilon_r^R. \tag{4.24}$$

$$\overline{u}_\lambda^\Lambda - \underline{u}_{t,\upsilon}^T \leq M(f_{t,\lambda,\upsilon} + g_{t,\lambda,\upsilon}) \qquad \forall t \in T, \lambda \in \Lambda : \upsilon \in \widehat{\Upsilon}_\lambda^\Lambda \cap \bigcup_{r \in R_t} \Upsilon_r^R. \tag{4.25}$$

$$\underline{u}^T_{t,v} - \overline{u}^\Lambda_\lambda < M(1 + f_{t,\lambda,v} - g_{t,\lambda,v}) \qquad \forall t \in T, \lambda \in \Lambda : v \in \widehat{\Upsilon}^\Lambda_\lambda \cap \bigcup_{r \in R_t} \Upsilon^R_r. \tag{4.26}$$

$$m_{t,v} \geq g_{t,\lambda,v} \qquad \forall t \in T, \lambda \in \Lambda : v \in \widehat{\Upsilon}^\Lambda_\lambda \cap \bigcup_{r \in R_t} \Upsilon^R_r. \tag{4.27}$$

$$m_{t,v} \leq \sum_{\substack{\lambda \in \Lambda: \\ v \in \widehat{\Upsilon}^\Lambda_\lambda}} g_{t,\lambda,v} \qquad \forall t \in T : v \in \bigcup_{\lambda \in \Lambda} \widehat{\Upsilon}^\Lambda_\lambda \cap \bigcup_{r \in R_t} \Upsilon^R_r. \tag{4.28}$$

We apply the same method to reduce the number of binary variables that was applied for RECIFE-TTP. This method, based on the one proposed by Pellegrini et al. (2015), exploits the topology of the railway infrastructure which allows to group a set of consecutive TDSs so they can be represented as a single binary variable in our formulation, instead of several. A detailed description of this method is given in Section 3.4.6. Besides of reducing the number of y variables, here we also achieve a great reduction of the number of z, f and g binary variables.

4.5 Solution Approach

In this section, we present the four algorithms that we propose and use for our experiments. We establish a *whole time limit* for each algorithm. This is the total wall clock time available for an execution. After this time, the best feasible solution (if any) is returned. Note that a solution is considered optimal only if the algorithm completes its optimality proof. Moreover, some of the proposed algorithms are two-phase algorithms, in these cases, a *first-phase time limit* is also established. More details about the behaviour of the algorithms once the first-phase time limit elapses are given in the respective descriptions.

RECIFE-MAINT: Full (RM-F)

This algorithm consists in solving the complete formulation described in Section 4.4 by using a MILP solver.

RECIFE-MAINT: Battery (RM-B)

This algorithm consists in solving a constrained version of RECIFE-MAINT by using a MILP solver. This constrained formulation is conceived to emulate the train batteries approach. Recall that the batteries approach, as reviewed in Section 4.3.1, consists in scheduling train batteries in both directions over a single track segment without paying attention to specific route of the trains in other portions of the infrastructure.

In a pre-processing step of RM-B, we assign a route to each train: First, we verify that the default route does not pass through some MAs' directly affected TDSs. If this is not the case, the default route is assigned to the train. Otherwise, one of the alternative routes of the concerned train is assigned instead. To this end, first, all alternative routes that are directly affected by a MA are discarded, then the route with the most similar TDSs sequence, with respect to the default one, is chosen. If no such route exists, the default route is maintained and assigned to the train, even if this implies that the train will have to wait until the MA is finished for completing its journey.

The constrained version of RECIFE-MAINT we use in this algorithm does not consider alternative routes, i.e., trains can only use their assigned route (either the default one or the new one assigned in the pre-processing step). More precisely, all binary variables related to routing (x) are suppressed from the formulation and the related constraints are updated accordingly.

The so obtained optimal solution can be interpreted as including train batteries, in which the values of the three parameters characterizing a schedule obtained by using the batteries approach (Section 4.3.1) are also optimal: the number of train batteries, the number of trains per battery and the direction of the first battery. For this reason we can assert that the optimal solution obtained by RM-B may also be interpreted as the result of an optimal battery approach. As such, RM-B is aimed to serve as a benchmark to compare the performance of the other algorithms with

the current practice.

RECIFE-MAINT: Battery-Full (RM-BF)

It is a two-phase algorithm where RM-B is used to obtain a feasible solution, then this solution is used to initialize RM-F.

After the first-phase time limit elapses, the best solution obtained by RM-B is used to initialize RM-F. If no feasible solution is found, the execution of the first-phase is extended until a feasible solution is found or the whole time limit expires. However, if before the first-phase time limit elapses the optimality of a solution is proven, the first phase ends and the remaining time is devoted to the second-phase.

RECIFE-MAINT: Greedy-Full (RM-GF)

RM-GF is a two-phase algorithm in which the first phase uses a heuristic algorithm that allows re-routing to generate an initial solution for RM-F.

The heuristic algorithm in the first phase is based on a greedy approach that plans trains one at the time. Once a train is planned, its route and schedule can not be changed. Algorithm 1 describes the main structure of this greedy algorithm which can be divided into three main activities:

- Timetable construction: A timetable is build by planning trains, one by one, following a predefined *train order*. Once a train is processed, its *plan* (chosen route and schedule) cannot be modified. Once all trains are planned in the timetable, the objective function value corresponding to the resulting timetable is calculated.

- Train planning: The best feasible schedule attainable with each route of the train is computed, and the corresponding objective function contribution value (*objCont*) is calculated. Then, the route with the lowest objCont is assigned to the train along with its respective schedule.

Algorithm `GreedyAlgorithm()`

while $currTime < TimeLim$ **do**
  $ord \leftarrow generateOrder(T)$
  $TT_{ord} \leftarrow$ $constructTimetable(ord)$
  $objVal_{ord} \leftarrow$ calc. $objVal$ of TT_{ord}
end
$bestTT \leftarrow TT_{ord'}$ with lowest $objVal_{ord'}$
return $bestTT$

Procedure `constructTimetable(`ord`)`

$TT \leftarrow$ empty timetable
foreach $t \in ord$ **do**
  $plan_t \leftarrow planTrain(TT, t)$
  $TT \leftarrow add(plan_t)$
end
return TT

Procedure `planTrain(`TT, t`)`

foreach $r \in R_t$ **do**
  $sched_r \leftarrow scheduleRoute(TT, r)$
  $objCont_r \leftarrow$ calc. $objCont$ of $sched_r$
end
$plan_t \leftarrow r', sched_{r'}$ with lowest $objCont_{r'}$
return $plan_t$

Procedure `scheduleRoute(`TT, r`)`

$sched_r \leftarrow$ empty schedule
for $i \leftarrow 1$ **to** $\left|\Upsilon_r^R\right|$ **do**
  $sched_r \leftarrow$ schedule $\upsilon_i \in \Upsilon_r^R$
  if $conflict(TT, sched_r)$ **then**
    $i \leftarrow$ $solveConflict(TT, sched_r)$
  end
end
return $sched_r$

Algorithm 1: Greedy Algorithm.

- Route scheduling: Given a route, the schedule is generated in a sequential way, i.e., the utilisation of each TDS in the route is scheduled progressively, from the first to the last one. When a conflict arises, i.e., a superposition of a TDS's utilization time with a MA or another previously planned train, a longer stay is added to the closest previous TDS where the train is allowed to experience delay, and the scheduling is resumed starting from said TDS. The scheduling process is finished once the utilisation of the last TDS of the route is set.

Note that the train order may have a high impact on the objective function value of the generated timetable. Indeed, some trains, when planned early in the process, can cause heavy delays to the trains planned later. To mitigate the effect of

the train order, we include in the greedy algorithm a mechanism to explore several possibilities. A timetable is generated with each train order, and the one with the lowest objective function value is chosen as the solution of the greedy algorithm. The number of train orders explored depends on the first-phase time limit. Indeed, as long as the first-phase lasts, new random train orders are automatically generated. However, we also define a set of train orders that are considered by the algorithm before any random order. The train orders we propose are described below in the same sequence as they are considered by the algorithm:

- Trains with longer routes are planned first: These trains potentially have more conflicts with the rest because of the length of their routes. Hence, the total delay experienced by these trains might be reduced if planned first.

- Trains with shorter routes are planned first: Ideally, by planning these trains first, the number of conflicts among them will remain low. Hence, probably a significant number of trains are planned without any delay before trains with longer routes are planned.

- Trains with fewer stops are planned last: Due to the green wave policy that must be respected by OTs, these trains have a limited number of locations where a longer stay may be introduced. Moreover, the objective function sums the accumulated delay at every stop of each train. Hence, trains with fewer stops potentially contribute a lower value to the objective function.

- Trains with more stops are planned last: These trains have more possible locations where a longer stay may be introduced. Therefore, these trains are more flexible for rescheduling. Ideally, the longer stays are only applied, if necessary, to its latest stops before the conflict.

Two additional considerations are applied when generating a new train order: First, MTs are always planned before any other train. This is done because of the hard timing constraints related to MTs, i.e., MTs must arrive at, and depart from, the MA location at the exact same time as the MA starts and ends, respectively, as

defined in Constraints (4.17) and (4.18) in Section 4.4. Second, trains with rolling stock re-utilisation are planned right after MTs. This is done to avoid deadlock situations in a straightforward way.

Typically, several orders are considered during the first-phase time interval, but if the first-phase time limit elapses before the first timetable is obtained, the execution of this phase is extended until a solution is generated or the whole time limit expires.

4.6 Experiments & Results

The algorithms were implemented in C++ using the Concert Technology for the IBM ILOG CPLEX v12.6 MILP solver (IBM, 2015) and the RCM-LIB (Arenas et al., 2016).

We set up two rounds of experiments using the case study described in the Section 4.6.1. The objectives of the first round of experiments are: Initially, to evaluate the capacity of RECIFE-MAINT and the proposed algorithms to obtain solutions to instances of the problem with different characteristics. Then, to assess the solutions improvement with respect to the solutions applied in current practice, i.e., the solutions obtained by RM-B. Finally, to obtain insights about the features of difficult instances and the capability of the algorithms to deal with them. The second round is designed to test the performance of the algorithms when dealing with very large instances.

The experiments are performed in a computer with eight Intel Xeon 3.5 Ghz processors and 128 GB RAM. The whole time limit is set to 1 hour for all algorithms. The first-phase time limit is set to 15 minutes for the two-phase algorithms (RM-BF and RM-GF). These time limits are the same for both rounds of experiments.

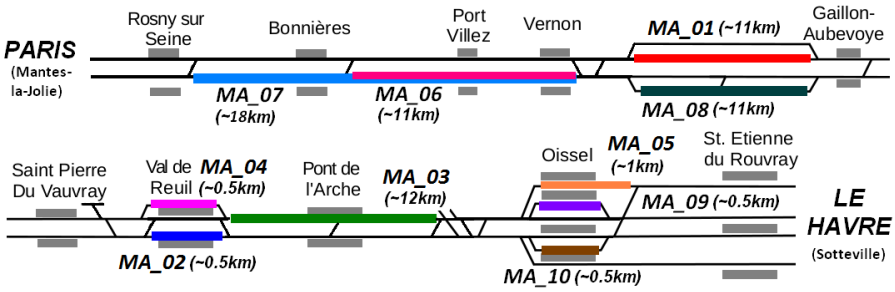


Figure 4.3: Section of the Paris - Le Havre line containing MAs.

4.6.1 Case Study description

We use the same microscopic representation of the infrastructure of the Paris - Le Havre corridor introduced in Section 3.6.1. For this control zone we define a set of ten unplanned MAs to be performed (one per instance) by mimicking some MAs actually performed on this corridor in 2012. The *locations* of these MAs are shown in Figure 4.3. When a MA is performed, all the adjacent tracks are subject to a TSL of 40 km/h, the maximum speed allowed in the nominal situation being 160 km/h.

Each MA requires the presence of one MT to be performed. The MTs must traverse a route connecting the location of the concerned MA and a shunting yard. We consider two shunting yards, one at Mantes-la-Jolie, between Rosny sur Seine and Paris, and one at Sotteville, between St. Etienne du Rouvray and Le Havre. For MA_01, MA_03, MA_04, MA_05 and MA_09, the inbound MTs depart from Mantes-la-Jolie and the outbound MTs arrive to Sotteville. For the other MAs, the inbound MTs depart from Sotteville and the outbound MTs arrive to Mantes-la-Jolie.

4.6.2 First round of Experiments

4.6.2.1 Instances

Each instance is defined by a combination of: a MA location, a MA duration and a time horizon. The ten MA locations are described in Section 4.6.1 and shown in

Figure 4.3. Three *durations* are considered for each MA location: 60, 90 and 120 minutes. We consider time horizons of four hours, starting at randomly selected times. The MA is planned in the middle of the time horizon. Ten different time horizons are randomly drawn for each couple MA location - MA duration, i.e., we tackle 300 instances in total.

	Ins.	Tra.	DIRiT	TSLiT	Cont.Vars	Bin.Vars	Constrs
all	300	48	5	5	158 K	6 K	344 K
MA location							
MA_01	30	44	6	6	153 K	5 K	330 K
MA_02	30	48	6	2	162 K	6 K	346 K
MA_03	30	48	8	10	157 K	6 K	353 K
MA_04	30	50	6	1	161 K	6 K	341 K
MA_05	30	46	6	4	153 K	5 K	322 K
MA_06	30	45	6	8	155 K	5 K	352 K
MA_07	30	47	7	8	159 K	6 K	376 K
MA_08	30	48	0	6	161 K	6 K	349 K
MA_09	30	48	3	8	162 K	6 K	345 K
MA_10	30	47	3	5	157 K	5 K	325 K

Table 4.2: Characteristics of the Instances: First round of experiments. (Ins.: Number of instances, Tra.: Median number of trains, DIRiT: Median number of directly impacted trains, TSLiT: Median number of TSL impacted trains, Cont.Vars: Median number of continuous variables, Bin.Vars: Median number of binary variables, Constrs: Median number of constraints.)

Table 4.2 resumes the main characteristics of the tackled instances. The first part concerns all 300 instances while the second part classes the instances depending on the MA location. We report the number of instances, the median number of trains present, the median number of trains impacted by the MA, the median number of continuous and binary variables and the median number of constraints in RECIFE-MAINT. Indeed, there are two ways in which MAs can impact trains in the existing

timetable:

- Directly impacted Trains (DIRiT): Trains in this category have their default route directly affected by a MA, i.e., the trains are scheduled to use a track segment while the MA is performed on the track segment itself. Trains in this category need to be either re-routed or delayed until the MA finishes.

- TSL impacted Trains (TSLiT): Trains in this category have their default route directly affected by a TSL due to a MA, i.e., the trains are scheduled to use a track segment during a TSL period due to a MA.

4.6.2.2 Results

The results of the first round of experiments are reported in Table 4.3. The same instance classifications as described for Table 4.2 are used: all instances and instances classed by MA location. This table reports the results of the algorithms in terms of three performance indicators: the ability to find feasible solutions, the ability to find and prove the optimality of the solutions and the improvements to the objective function value with respect to the emulation of the current practice (RM-B). In particular, each column reports:

- % Feasible Sol.: The percentage of instances where a feasible solution is obtained by the algorithms.

- % Optimal Sol.: The percentage of instances where the optimal solution is found and proven by the algorithms .

- Sol. Improv.^{a} (%) and Sol. Improv.^{b} (%): The mean percentage of improvement to the objective function value with respect to RM-B, for instances where all algorithms find a feasible solution and for instances where RM-F does not manage to do so, respectively.

We indicate in bold the best performance across the algorithms according to the solution improvement indicator.

Class	% Feasible Sol.			% Optimal Sol.			Sol. Imp.[a] (%)			Sol. Imp.[b] (%)	
	RM-F	RM-BF	RM-GF	RM-F	RM-BF	RM-GF	RM-F	RM-BF	RM-GF	RM-BF	RM-GF
all	87	100	100	61	68	68	-46	28	**40**	-11	**34**
MA location											
MA_01	100	100	100	50	60	57	-4	**16**	**16**	-	-
MA_02	100	100	100	97	100	100	**91**	**91**	**91**	-	-
MA_03	50	100	100	17	20	20	-48	12	**27**	-6	**25**
MA_04	100	100	100	73	97	93	44	**49**	**49**	-	-
MA_05	100	100	100	97	97	100	**69**	**69**	**69**	-	-
MA_06	77	100	100	17	13	23	-39	22	**44**	-14	**32**
MA_07	50	100	100	3	7	10	18	25	**36**	-14	**42**
MA_08	97	100	100	67	83	80	-990	**2**	1	**0**	**0**
MA_09	100	100	100	87	100	100	53	**58**	**58**	-	-
MA_10	100	100	100	100	100	100	**59**	**59**	**59**	-	-

Table 4.3: Results of the first round of experiments using RM-F, RM-BF, and RM-GF. (% Feasible Sol.: Percentage of instances where at least a feasible solution is obtained, % Optimal Sol.: Percentage of instances where the optimal solution is found and proven, Sol. Imp. (%): Mean percentage of improvement to the objective function value with respect to the solution obtained by RM-B, [a] for instances where all algorithms find a feasible solution, [b] for instances where RM-F does not find a feasible solution). Bold indicates the best performance across the algorithms according to the solution improvement indicator.

First of all, we remark that both RM-BF and RM-GF obtain at least a feasible solution for all instances within the whole time limit fixed. Instead, RM-F fails to produce a feasible solution for 13% of the instances. More specifically, as shown in the lower part of the table, RM-F has more difficulty to deal with MA_03 and MA_07 instances, where a feasible solution is found for only 50% of the instances. Even with a higher percentage of success, MA_06 and MA_08 also appear to be very problematic for RM-F.

In general, the number of optimal solutions proven by the three algorithms is similar, although a slightly better performance is reported for RM-BF and RM-GF (68% of all instances versus the 61% of RM-F). We remark as well that, regardless

of the used algorithm, a distinctly low percentage of optimal solutions was found for MA_03, MA_06 and MA_07 instances. In particular, in the best case 10% of the solutions for MA_07 instances were proven optimal by the RM-GF algorithm. These results give an initial hint concerning the different difficulties of the tackled instances.

Concerning the performance of the algorithms in terms of solution improvement with respect to RM-B, we compare separately the instances where all three algorithms find at least a feasible solution and those instances where RM-F does not. Recall that RM-B is considered as a reference as it is a better performing proxy of the current practice, as explained in Section 4.5.

Consider first the 87% of instances where all three algorithms find at least a feasible solution. The solutions returned by RM-BF and RM-GF are better than the ones found by RM-B, in mean of 28% and 40%, respectively. Instead, the solutions obtained by RM-F are in mean 46% worse than the ones produced by RM-B.

By closely analysing the results of RM-F, we observe that this algorithm is systematically outperformed by both RM-BF and RM-GF, except for those instances where all three algorithms find the optimal solution, either proving the optimality or not. This happens in 179 instances. The performance difference is especially marked in MA_03, MA_06 and MA_08 instances, where RM-F obtained solutions considerably worse than RM-B. Note that for MA_08 instances, the mean value of solution improvement is particularly low (-990%). From a detailed exam, it emerges that this is due to one specific instance where RM-F obtains a solution 14000% worse than the one obtained by RM-B: the first solution has a total arrival time delay of 1164560 seconds, the second of 8299 seconds, for a MA duration of 120 minutes. By not considering this outlier, the mean value of solution improvement of RM-F for MA_08 instances raises to -42%. The overall low performance of RM-F with respect to the two-phased algorithms (RM-BF and RM-GF) shows that, as expected, it is very useful to provide an initial solution to the MILP solver to obtain satisfactory results in a reasonable amount of time.

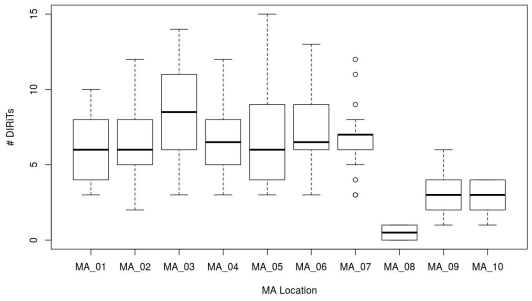
By observing the differences in performance between RM-BF and RM-GF, we remark that for most instances groups both algorithms have a very similar mean improvement value. Yet, for MA_03, MA_06 and MA_07 instances, the solutions obtained by RM-GF are better than the ones obtained by RM-BF. Indeed, we observe that the overall solution improvement of RM-GF is in mean 12% better than RM-BF.

Consider now the 13% of instances where RM-F cannot find a solution within the time limit. RM-GF again outperforms RM-BF. This time the difference between the improvements over RM-B achieved by these algorithms is in mean 45%. Furthermore, RM-BF obtains, in mean, worse solutions than the ones returned by RM-B. This is due to the existence of a first-phase time limit. Recall that RM-BF uses the best solution obtained by RM-B within the first-phase time limit (15 minutes) to initialize the MILP solver for the complete formulation. Therefore, whenever RM-BF is outperformed by RM-B is because the second phase of RM-BF starts with a sub optimal solution of RM-B. In the 27 instances (9%) in which this is the case, RM-B uses the remaining time more efficiently than RM-BF, which suffers for the very large search space.

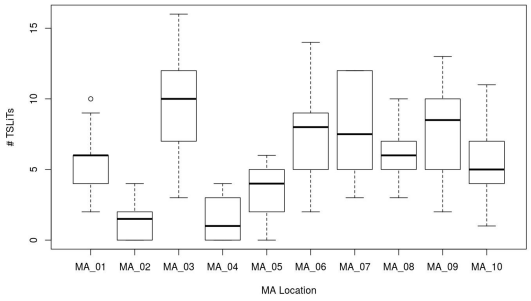
We perform several Wilcoxon rank-sum tests. We obtain that for instances where all three algorithms find a feasible solution, the solutions of all algorithms are significantly better than those obtained by RM-B with a confidence level of 0.95. Moreover, we can also state that, for all instances, the solution improvements made by RM-GF are significantly higher than the improvements made by RM-BF with a confidence level of 0.95. Hence, we conclude that the proposed two-phase algorithms are indeed applicable, since they always provide solutions to the real case study. Moreover, we can state that they are well performing, since they outperform the optimized version of the approach used in the current practice.

By analysing the number of impacted trains, i.e., DIRiTs and TSLiT, and comparing these with the GAP of the solutions produced by RM-GF, our best performing algorithm, we can obtain some insights on the characteristics of instances which result difficult. Recall from Section 3.6.3 that we consider the GAP as an indicator of

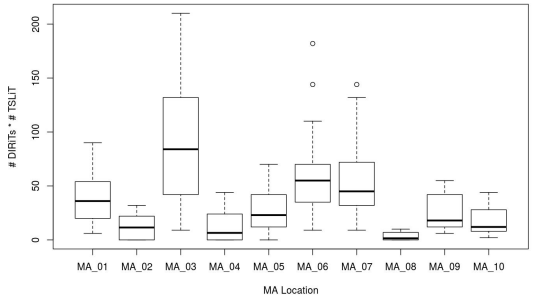
the difficulty of an instance for the algorithm, i.e., the higher the GAP, the more difficult the instance. Figure 4.4 presents four boxplots showing, for all instances classed by MA location, the distributions of: The number of DIRiTs (Figure 4.4a), the number of TSLiTs (Figure 4.4b), the product of DIRiTs and TSLiTs (Figure 4.4c) and the GAP of solutions obtained by RM-GF (Figure 4.4d). Each box represents the distribution of the observations corresponding to the 30 instances of each MA location. The horizontal line within the box represents the median of the distribution, while the extremes of the box represent the first and third quartiles, respectively; the whiskers show the smallest and the largest non-outliers in the data-set and dots correspond to the outliers. Note that the medians of the distributions in Figures 4.4a and 4.4b are also reported in Table 4.2.



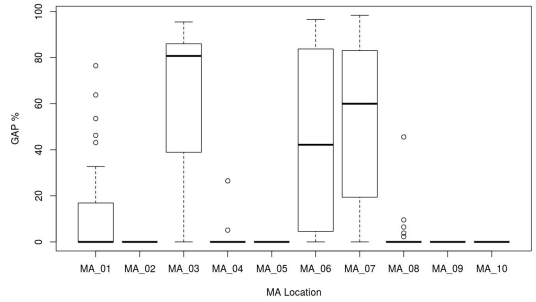
(a) DIRiT count by MA Location.



(b) TSLiT count by MA Location.



(c) DIRiT · TSLiT by MA Location.



(d) RM-GF's GAP by MA Location.

Figure 4.4: Boxplots representing the distributions of DIRiT, TSLiT, their product and the GAP of RM-GF solutions for all instances classed by their MA location.

By observing the distributions of DIRiTs in Figure 4.4a, we remark that MA_01 to MA_07 instances have a similar number of DIRiTs, with a median between 6 and 9, whereas MA_08 to MA_10 instances have a lower count of DIRiTs. By comparing these distributions with the distributions of the GAPs obtained by RM-GF (Figure 4.4d) it is not evident to establish a clear relation between the number of DIRiTs and the difficulty of the instances. For example, we observe that MA_01 and MA_05 instances have a similar distribution of DIRiTs. However, MA_01 instances are more difficult to solve: for 13 out of 30 MA_01 instances the GAP is strictly greater than the median value 0, whereas the solutions for all MA_05 instances have a GAP of 0.

Although the relation between the TSLiTs count and the difficulty of the instances seems slightly more perceptible, as we can observe by comparing the distributions shown in Figures 4.4b and 4.4d, this relation is not always clear. For example, MA_06 and MA_09 instances have a similar distribution of TSLiTs, but MA_09 instances are significantly easier, as 100% of the solutions are proven to be optimal.

When considered independently, neither DIRiTs nor TSLiTs are able to clearly express the difficulty of an instance. However, when coupled by multiplying their values, a more apparent relation emerges. Figure 4.4c shows the distribution corresponding to this product. We observe that the distributions with higher product of DIRiTs and TSLiTs indeed correspond to the most difficult groups of instances shown in Figure 4.4d: MA_03, MA_07, MA_06 and MA_01, in descending order. This is corroborated by the high value of the correlation coefficient of 0.78, which gives the quality of a least squares fitting to the product and GAP distributions (Edwards, 1976). The correlation coefficient computed for DIRiTs and TSLiTs is 0.47 and 0.62, respectively.

4.6.3 Second Round of Experiments

4.6.3.1 Instances

In the second round of experiments, the size of the time horizon is set to 8 hours. The duration of MAs is set to 4 hours, which corresponds to one of the largest MA durations commonly allowed during daytime operations on the control zone under study. Five time horizons are randomly set for each MA location, hence, we tackle a total of 50 instances.

	Ins.	Tra.	DIRiT	TSLiT	Cont.Vars	Bin.Vars	Constrs
all	50	98	15	14	324 K	21 K	844 K
MA location							
MA_01	5	102	18	18	326 K	22 K	872 K
MA_02	5	101	15	2	324 K	21 K	835 K
MA_03	5	98	20	24	332 K	22 K	899 K
MA_04	5	99	16	3	336 K	21 K	851 K
MA_05	5	103	17	9	351 K	22 K	892 K
MA_06	5	95	15	15	310 K	18 K	819 K
MA_07	5	89	15	18	321 K	20 K	894 K
MA_08	5	88	1	12	310 K	20 K	804 K
MA_09	5	98	8	18	314 K	21 K	803 K
MA_10	5	94	7	10	313 K	19 K	771 K

Table 4.4: Characteristics of the Instances: Second round of experiments. (Ins.: Number of instances, Tra.: Median number of trains, DIRiT: Median number of directly impacted trains, TSLiT: Median number of TSL impacted trains, Cont.Vars: Median number of continuous variables, Bin.Vars: Median number of binary variables, Constrs: Median number of constraints.)

Table 4.4 reports the main characteristics of these large instances. The first part of the table concerns all 50 instances while the second part classes the instances depending on the MA location. As we did in Table 4.2 for the first round of ex-

periments, for each instance class, we report the number of instances, the median number of trains present, the median number of trains impacted by the MA, the median number of continuous and binary variables and the median number of constraints in RECIFE-MAINT.

4.6.3.2 Results

The results are resumed in Table 4.5 in the same form as in Table 4.3. Note that we do not report the results for RM-F. The reason for this is that RM-F is able to find a feasible solution for only 20% of the instances (10 out of 50) within the whole time limit.

First, we observe that both algorithms, RM-BF and RM-GF, obtain a feasible solution for all instances. Nonetheless, the mean number of optimal solutions proven is considerably reduced with respect to the first round of experiments. Indeed, as the time horizon and MA duration grow larger, the median number of trains, DIRiTs and TSLiTs increase as well, between two and three times with respect to the first round. As postulated in the discussion of the results of the first round of experiments, the product of these is an effective indicator of the difficulty of the instances. This relation is mildly corroborated in this round: we performed a similar analysis as the one described in Section 4.6.2.2, obtaining the correlation coefficients of 0.27, 0.48 and 0.51 for DIRiTs, TSLiTs and the product of them, respectively, with regard to the GAP obtained by RM-GF. The highest coefficient again is the one linking the GAP to the product of the two numbers of affected TDSs. The lower values of these correlation coefficients can be explained because of the contribution of another factor that undoubtedly increases the difficulty of the instances, that is, the size of the formulation. Indeed the size of an instance's formulation has an impact in the difficulty of solving the instance itself. By observing the size of the formulations in the first round of experiments (Table 4.2) and the percentage of optimal solutions obtained (Table 4.3) we notice that the size of the formulations are always manageable. Hence, the impact of the different indicators is quite evident

Class	% Feasible Sol.		% Optimal Sol.		Sol. Imp. (%)	
	RM-BF	RM-GF	RM-BF	RM-GF	RM-BF	RM-GF
all	100	100	12	10	-115	**29**
MA location						
MA_01	100	100	0	0	**3**	-28
MA_02	100	100	0	0	10	**89**
MA_03	100	100	0	0	-92	**27**
MA_04	100	100	0	0	**9**	4
MA_05	100	100	20	20	18	**49**
MA_06	100	100	0	0	-2	**18**
MA_07	100	100	0	0	-200	**38**
MA_08	100	100	0	0	**0**	-16
MA_09	100	100	40	40	**35**	16
MA_10	100	100	60	40	68	**71**

Table 4.5: Results of the second round of experiments using RM-BF, and RM-GF. (% Feasible Sol.: Percentage of instances where at least a feasible solution is obtained, % Optimal Sol.: Percentage of instances where the optimal solution is found and proven, Sol. Imp. (%): Mean percentage of improvement to the objective function value with respect to the solution obtained by RM-B.) Bold indicates the best performance across the algorithms according to the solution improvement indicator.

in terms of the correlation coefficient. Instead, as shown in Tables 4.4 and 4.5, the size of the instances' formulations in the second round of experiments is almost tripled and the percentage of optimal solutions is considerably lower than in the first round. Here, apparently, the formulations become too large to be managed within the whole-time limit and the solution space is not efficiently explored, disregard of the difficulty due to the affected TDSs.

Concerning the performance of the algorithms, we notice that the overall performance of RM-BF, in terms of solution improvement with respect to RM-B, is greatly reduced in this round. This is particularly true when dealing with the most difficult

instances of MA_03, MA_06 and MA_07, which have the highest product of DIRiTs and TSLiTs. Note that no outliers are present among the solutions.

Despite of the increased size and difficulty of the instances in this round, the mean percentage of solution improvement obtained by RM-GF is 29%, which remains close to the one obtained in the first round. Moreover, RM-GF outperforms RM-BF in this round as well, the difference being in mean 144%. This outcome is confirmed by the Wilcoxon rank-sum test, in which we observe that the results obtained by RM-GF are significantly better than the results of both RM-B and RM-BF with a confidence level of 0.95. However, we remark that for MA_01 and MA_08 instances, the mean percentage improvement of RM-GF with respect to RM-B is negative.

By comparing the mean solution improvement obtained by the algorithms for MA_01 and MA_03 instances, we observe that RM-BF outperforms RM-GF for MA_01 instances but is greatly outperformed for MA_03 instances. Note that the product of the medians of DIRiTs and TSLiTs for MA_01 and MA_03 instances are 324 and 480, respectively, which means that MA_03 instances are, in principle, more difficult to solve than those of MA_01. Indeed, MA_03 instances are more difficult to solve than MA_01 for RM-BF, as the median GAP are 99.8% and 93.3% respectively. However, RM-GF have more difficulties to solve the easier MA_01 instances as for these it obtains a median GAP of 99.8%, compared to 97.8% for those of MA_03. This may mean that RM-GF has difficulties to deal with some particularities of MA_01 instances.

To understand this difficulty we thoroughly examined the MA_01 instances: we analysed their microscopic characteristics and we compared the timetables produced by the greedy component of RM-GF with those obtained by RM-B. By doing so, we identified two main issues, both related to the train orders used by the algorithm. Recall that although several orders are considered by the greedy algorithm (Section 4.5), they all start with the MTs and immediately after, the trains with rolling stock re-utilisation.

The first issue is explained with an example of one particular instance. In this instance, we find that a train with a rolling stock re-utilisation operation at the Gaillon Aubevoye station, let it be t, arrives at this station with a delay due to the TSL of MA_01. Another train t', which is initially scheduled to pass trough the Gaillon Aubevoye station just before t and has no alternative routes, is planned after t because of the imposed order, therefore t' is delayed for at least the entire duration of the rolling stock re-utilisation operation. The delay imposed to t', in turn, cause more delays to other trains planned afterwards, thus originating a timetable with severe overall delay. We discover that when this same instance is solved by RM-B, t is actually delayed elsewhere, allowing t' to pass trough Gaillon Aubevoye first, thus avoiding the multiple delays occasioned to other trains. In this specific instance, t' will always be delayed by t because of the order imposition of the greedy algorithm. This same circumstance may occur in more instances with other trains having rolling stock re-utilisation or MTs, and cannot be avoided due to the fact that these two types of trains are always planned first in the current version of the greedy algorithm.

The second issue is less evident since it is given by a circular dependency relation between three or more trains that is neither known *a priori* nor easy to identify. Consider the example given in Figure 4.5: trains t, t' and t'' present a circular dependency between them such as delaying t causes a delay on t', delaying t' causes a delay on t'' and delaying t'' causes a delay on t. Consider as well that delaying one of these trains, might also cause delays to other trains outside this circular dependency. Ideally, no delay should be added to the trains within the circular dependency, but if necessary, the best strategy might be to add a small amount of delay to each one of them to maintain the balance. However, this is not how the core of the greedy algorithm works, as one train is planned at each time regardless of the potential conflicts it may cause with the trains planned afterwards. Taking this into account, another strategy could be to identify the most crucial train among those in the circular dependency, e.g., the one that causes more delay to other trains outside the circular dependency, and place it in the train order always before the others. But the current version of the greedy algorithm does not consider train dependencies when

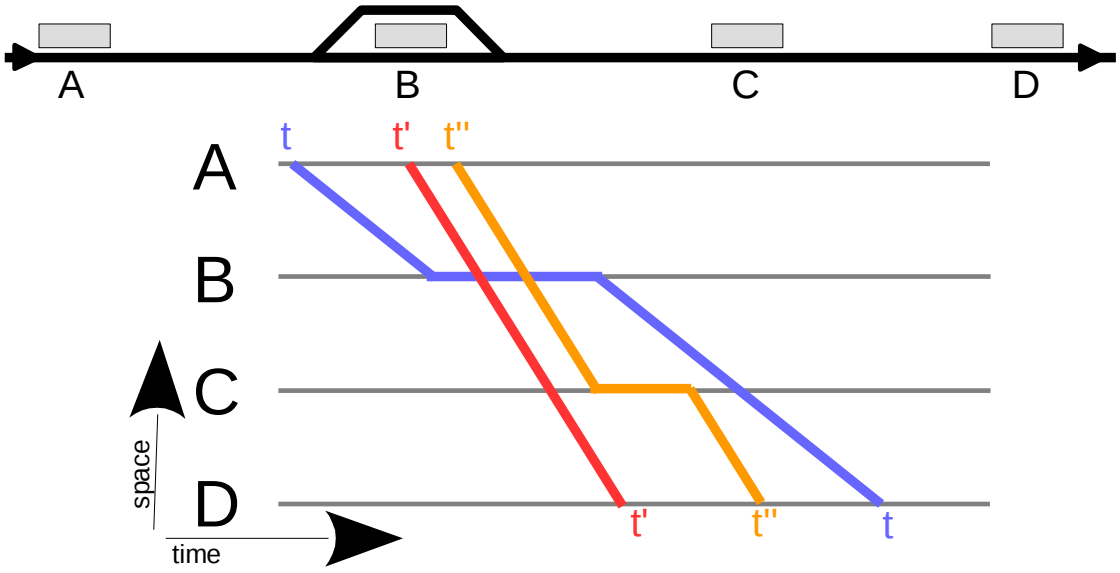


Figure 4.5: Example of circular dependency between trains. Three trains t, t' and t'' run between stations A and D. t is overtaken by t' and t'' in station B, the only available overtaking location. First, if the departure of t from A is delayed, it causes a delay on t'. Next, by delaying t', t'' is delayed as well. Finally, because t must be overtaken by t'' in B, and t'' has a scheduled stop in C, whenever t'' is delayed it causes a delay on t.

generating the train orders, and although random orders are also generated, these do not guarantee that an adequate order is proposed. These shortcomings clearly affect the overall performance of RM-GF for some groups of instances, and this effect is increased because of the size of the instances in this round of experiments. By assessing the improvements achieved by the second phase of the algorithms for all groups of instances, we discover that for MA_01 instances, the mean improvement achieved by the second phase is only 3% for RM-BF and RM-GF. This evidences the importance of carefully designing the first-phase algorithm, as it can play a critical role in the final outcome of the complete algorithm. However, this does not mean that we can diminish the role of the second phase of the algorithms, since important improvements are obtained, e.g., a mean improvement of 57% is achieved by the second phase of RM-BF for MA_10 instances. The mean solution improvements obtained by the second phase of RM-BF and RM-GF for all instances in this round

of experiments are 14% and 19% respectively.

4.7 Conclusion

In this chapter, we discussed the problem of timetable rearrangement to cope with infrastructure MAs. This is a relevant problem because the realization of unplanned MAs may severely impact the scheduled train circulations in the railway infrastructure. Moreover, the current context of increasing demand of railway capacity requires an approach to solve this problem while exploiting the capacity as efficiently as possible.

There is a limited number of approaches proposed in the literature to deal with this specific problem. All of them are based on macroscopic representations of the infrastructure, that may overestimate or underestimate the actual capacity of the infrastructure, leading to the practical infeasibility of the timetables or the sub-optimal exploitation of the capacity, respectively. Moreover, these approaches often disregard specific aspects strongly related with the performance of MAs, such as the presence of MTs and the imposition of TSLs. Indeed, the impact of TSLs is often underestimated, in the results of our experiments we found out that in several cases the total running time of some trains is considerable increased, which not only alters the initial schedule of said trains but also may generate a number of conflicts with others.

In this context, we presented RECIFE-MAINT: a MILP formulation to solve the problem of rearranging a timetable to cope with MAs while considering the planning of MTs and imposition of TSLs. RECIFE-MAINT uses a microscopic representation of the infrastructure and guarantees the feasibility of the produced timetables while optimizing the railway capacity utilization.

We proposed three algorithms implementing RECIFE-MAINT, which we tested on a real case study in France. Additionally, we considered an algorithm that emu-

lates the current practice, which we used as a benchmark to compare the performance of the algorithms we proposed.

The results showed that our algorithms were able to produce rearranged feasible timetables in a reasonably short amount of time, even when dealing with very large instances. Moreover, the resulting timetables were significantly better, in terms of overall delay reduction, than the timetables obtained by applying the algorithm emulating the current practice.

Additionally, the analysis of the results allowed us to make conjectures about the characteristics of the instances which impact their difficulty. Indeed, the difficulty of an instance can not be precisely estimated by independently observing the number of trains whose default route is affected directly by either a MA or by a TSL of a MA. Instead, by considering the product of these values we obtain a high correlation coefficient with respect to the GAP distributions of the solutions, which we consider as an indicator of the difficulty of an instance.

Perspectives of future work include the reduction of the computational time to solve larger instances of the tackled problem. Indeed, the results showed that by using a two-phase algorithm the computational time is significantly reduced. However, the results also showed that the first phase should be carefully designed to deal with specific characteristics of some instances, that may heavily affect the performance of the whole algorithm. First, an analysis of these characteristics should be performed to identify the actually relevant ones. Then, an appropriate algorithm should be designed.

Further research perspectives may include the planning and scheduling of MAs in the timetable rearrangement process. Indeed, in the problem tackled in this chapter, the timetable is adapted to a fixed schedule of MAs. However, by allowing MAs' rescheduling, new possibilities may become available. For instance, it may be possible to schedule MAs when they do not impact traffic.

Chapter 5

Conclusion

5.1 Summary

In this thesis we presented our contributions towards improving the efficiency in which the railway capacity is exploited. More specifically, we focused on the elaboration and rearrangement of train timetables in nominal conditions and during the performance of MAs.

We started this thesis by offering a review of the main concepts and procedures of the railway system in Chapter 2. In particular, we defined the concept of capacity and presented the railway service planning process, drawing special attention to timetabling and maintenance planning activities. We presented as well how these procedures are carried out in the French railway system. The final part of the review concerns the two main approaches for representing the railway infrastructure: macroscopic and microscopic. We comprehensively presented both approaches and compared them to identify their strengths and drawbacks when used for (re)timetabling purposes.

Indeed, most optimization tools currently used by the IMs and RUs to process timetables are based on macroscopic representations of the infrastructure. The abil-

ity to deal with very large instances of the problem in a relatively short time and the small amount of data necessary to model the infrastructure are the main advantages of macroscopic approaches. However, because of the abstractions used, overestimations of the capacity utilization are often made to ensure the feasibility of the produced timetables. This leads to a possible inefficient exploitation of the railway capacity.

Microscopic approaches, instead, are able to precisely model several aspects of the actual railway system such as precise train movements and control command devices as the signalling and the interlocking systems. This allows the calculations to be highly accurate, which not only ensures the actual feasibility of the timetables, but also promotes an efficient exploitation of the railway capacity. Because of the larger amount of data and the longer time often required to execute microscopic approaches, they are typically only used on specific parts of a railway network, as main stations or complex junctions.

Considering the current context of increasing demand for capacity experienced by most railway systems, we believe that a microscopic approach must be used more extensively in the timetabling process. To this end, we introduced RECIFE-TTP, a microscopic formulation for the TTP. This formulation is thoroughly presented in Chapter 3 after a detailed description of the TTP and a review of the most relevant approaches which exist in the literature to tackle this problem.

Furthermore, as discussed in Chapter 2, the performance of MAs consumes an important part of the available capacity in the network. Although the plan of MAs is considered during the elaboration of timetables, the fact that the majority of timetabling approaches are based on macroscopic representations of the infrastructure implies that the produced timetables may be infeasible in practice or inefficient in terms of capacity exploitation. This is particularly true when the MAs require MTs or imply TSLs on adjacent tracks. In addition, unplanned MAs may be required at any time due to accidents or device malfunctions. In both cases, a rearrangement of the timetable may be necessary to effectively cope with MAs.

To tackle the problem of timetable rearrangement to cope with MAs we presented RECIFE-MAINT in Chapter 4. This microscopic formulation considers the imposition TSLs and the planning of MTs. Both of these aspects are neglected in the limited literature available dealing with this particular problem.

5.2 Conclusions

In accordance with the two main objectives of this thesis (O1 and O2) defined in the Introduction (Chapter 1), we proposed effective microscopic approaches to produce timetables and to rearrange them to cope with maintenance activities.

The series of experiments performed and reported in Chapters 3 and 4 have shown the practical applicability of our microscopic approaches (O1.a and O2.a). In particular, we could tackle instances of the problems considering control zones significantly larger than single stations or junctions. The best algorithms implementing RECIFE-TTP and RECIFE-MAINT were able to find a feasible solution for all the treated instances. In a high percentage of the cases, the optimality of the solution was actually proven.

Moreover, we confirmed the necessity of using a microscopic approach to ensure the feasibility of the timetables while maximizing the capacity utilization (O1.b). Indeed, the set of train-path requests used as an input for RECIFE-TTP proved to be microscopically infeasible. These train-path requests were provided by the French IM. They actually consisted in the timetable implemented on the considered portion of infrastructure, where a microscopic validation had not been performed. Concerning more specifically the results of the algorithms implementing the variants of RECIFE-TTP considered, we could point out that:

- Maintaining the structure of a train schedule by fixing its dwell times, which is often desired by the RUs, offers limited scheduling possibilities and may lead to inefficient schedules in terms of overall deviation from the train-path requests.

Instead, when the dwell times are allowed to be modified, provided that minimum values are always respected, better solutions can be found. Nevertheless, if the minimum dwell times do not contain any time buffer, it is possible for variable dwell time approaches to lose any dwell time buffer existing in the set of train-path requests. In this sense, timetables produced by fixed dwell time approaches may be more robust.

- Moreover, minimizing the time differences at every control point of the train's itineraries with respect to the desired schedules also allows the minimization of the impact of the modifications made to the trains' schedules at the border of the control zone considered. This allows the limitation of the impact on the whole network of a local optimization.

- Finally, by allowing multiple routes for the trains, more scheduling possibilities are considered. This in general leads to better solutions, even if considerably larger computational times are required. In our experiments, however, these larger computational times remained completely acceptable given the tactical nature of the TTP.

Concerning the timetable rearrangement problem to cope with MAs, we found that the limited number of approaches in the literature are all based on macroscopic representations of the infrastructure and neglect the imposition of TSLs and the presence of MTs. We confirmed in our experiments that the impact of TSLs should not be underrated and it has to be considered when concurrently dealing with MAs and train circulations. Indeed, we observed that the total running time of some trains is considerably increased when they are affected by TSLs. This implies important changes to the attainable schedules of the concerned trains, and this may produce a number of conflicts with other trains. We proposed three algorithms implementing RECIFE-MAINT, two of them based on a two-phase optimization process. Moreover, we proposed an algorithm that emulates the current practice, which we used as a benchmark. In general, timetables obtained by our two-phase algorithms were significantly better than the ones obtained by applying the algorithm emulating the

current practice (O2.b). However, even our best performing algorithm was unable to find optimal solutions for most of the very large instances tackled. We believe that further improvements can be achieved if a longer time is allocated to the algorithm. Recall that we set a computational time limit of one hour, which is quite short for a process that is to be carried out in a pre-operational level. We could draw two additional conclusions from the experiments reported:

- The proposed algorithms were able to find a feasible solution for most of the tackled instances in the available computational time. The two-phase algorithms produced better solutions than the single-phase one, and they always found at least a feasible solution. This shows that, as expected, by properly initializing the exploration of the search space, a microscopic approach can actually deal with large instances effectively.

- We identified some characteristics of the instances that have an impact on their difficulty. In particular, we remarked that the difficulty of an instance can not be clearly explained by regarding independently the number of trains whose default route is affected directly by either a MA or by a TSL of a MA. Instead, if the product of these values is considered, we obtain a high correlation coefficient with respect to the GAP distributions of the solutions, which we consider as an indicator of the difficulty of an instance.

In conclusion, we think that the work presented allows us to state the achievement of the objectives we had set. Nevertheless, it indeed rises several open issues for future research, which we discuss in the next section.

5.3 Perspectives

In this section we discuss several research perspectives that emerged during the performance of the works described in this thesis. We group these perspectives into five categories: Reduction of the computational time, (re)timetabling to cope with

maintenance, macroscopic-microscopic integration, timetable quality and timetable robustness.

The first category is related to the **reduction of the computational time**. Indeed, one of the main concerns of the use of microscopic approaches is the typically high computational times required to obtain the solution of a problem. In this thesis, we effectively reduced the computational time of our microscopic approaches by reducing the size of the instances and by implementing some specialized algorithms. More specifically, we reduced the number of binary variables by means of a boosting method introduced by Pellegrini et al. (2015) and we implemented a hybrid algorithm which sequentially combines a heuristic and a truncated exact method. Further ideas based in these two approaches are:

- *Reduction of the instance size.* In particular, the number of alternative routes of the trains have a large impact on the size of the instance. By considering fewer alternative routes, the size of the instance is reduced, which in turn can diminish the computational time to obtain a solution. Note that a crucial concern must be to reduce the search space without discarding the optimal solutions of the problem. Some research on this subject has already been made. For example, Samà et al. (2016) propose a metaheurisitc algorithm to reduce the alternative routes of trains. The infrastructure is microscopically represented and no particular attention is devoted to the fact that single stations or larger networks can be under consideration. Possibly, specializing algorithms as the ones by Samà et al. (2016) to take into account the nature of each portion of the considered infrastructure may allow an improvement of their performance for large instances.

- *Alternative algorithms.* A number of mathematical programming, heuristic and metaheuristic methods, may be used, and even combined to effectively reduce the computational time required by microscopic approaches. A non exhaustive list of these may include Lagrangian relaxation, Benders decomposition, column generation, tabu search, variable neighbourhood search, etc. First, a

thorough study is necessary to understand the compatibility of such methods with the microscopic problems treated in this thesis. Indeed, some methods may be more suitable than others for these problems. Next, extensive experimental analyses should be performed considering a variety of instances with distinct characteristics, e.g., infrastructure topologies, traffic types, etc.

The next category of perspectives concerns the **(re)timetabling to cope with maintenance**. In this category, we identify two extensions:

- *Maintenance Rescheduling.* The problem treated in Chapter 4 consists into adapting a timetable to a fixed schedule of MAs. By allowing the rescheduling of MAs, new timetabling possibilities become available, which may promote a more efficient exploitation of the capacity: MAs could be scheduled in time periods where they cause the minimum impact on the trains' circulations. Several scheduling constraints must be taken into account for MAs. For example, some MAs may be freely rescheduled during the time horizon, others may only be rescheduled into a small time window.

- *Maintenance planning.* As an extension of the previous point, the actual planning of MAs may be defined and optimized during the (re)timetabling process. This means that besides allowing the rescheduling of MAs, additional decisions are to be considered to optimize the planning of such activities. In this sense, the maintenance plan is represented by a set of modifiable MAs. These MAs are defined differently from what done in this thesis. Up to now they have been treated as individual and indivisible activities which are mainly defined by a location and a schedule. In this extension they are tasks whose locations, durations and schedules may be optimized. This implies the definition of a number of specific constraints concerning both the characteristics of each MA and the relations among them, e.g., allowed durations and locations, precedence and simultaneity relations, etc. For being practically relevant, such optimization will have to deal with much larger time horizons than the ones considered in this thesis. Therefore, adapted algorithms must be designed. For example,

they may implement the rolling horizon method to sequentially tackle smaller overlapping instances.

The next category of perspectives concerns the **macroscopic-microscopic integration**. In this thesis we effectively used microscopic approaches to tackle instances of the problem larger than single stations or junctions, which is currently their main role. This is true also in the majority of macro-micro integrated approaches. This enlargement of the scope of the microscopic component may have a twofold impact on the overall performance of the integrated approach. On the one hand, the number of iterations needed to converge into a feasible or optimal solution may decrease. Indeed, tackling fewer and larger microscopic instances may allow to quickly solve possible timetable incompatibilities. On the other hand, the computational time required for each microscopic iteration may become too large. To understand which impact is larger, a deep study should be performed.

The next category concerns the **quality of the timetables**. Two different solutions of the (re)timetabling problems tackled in this thesis may have the exact same objective value. However, if compared in a practical railway environment, considerable distinctions may become evident. For instance, consider a crowded intercity train with the exact same schedule in two different timetables, i.e., same objective value. The timetable in which the train is planned to stop at the widest platform of a station is better than the one in which the train is planned to stop at the narrowest platform of the same station. A number of other similar practical aspects may exist, which, if considered during the (re)timetabling procedure, can increase the quality of the produced timetables. A first step in this direction is to investigate and formalize the definition of these quality aspects that may be difficult to quantify. A second step is their integration in a microscopic approach. Note that this integration may increase the difficulty of the instances of the problem. Hence, a deep study concerning the trade-off between the increase of quality of the solutions and the increase of the difficulty of the instances should be performed.

The final category of perspectives concerns the **robustness of the timetables**.

The robustness may be considered as a quality aspect of timetables. However, we choose to separate it from the precedent point because of its high relevance in terms of operational feasibility of the timetables. Indeed, the robustness allows a timetable to absorb the small but ever-present perturbations of railway operations. To increase the robustness of the produced timetables, we identify at least two alternatives: the utilization of buffer times and the performance of stochastic optimization:

- *Buffer times.* Indeed, a straightforward method to increase the robustness of a timetable is the incorporation of buffer times to its train circulations. In principle, the larger the buffers are, the more robust a timetable becomes, but also, the less efficient becomes the capacity exploitation. To incorporate buffer times in the microscopic approaches tackling the problems discussed in this thesis, two steps should be made:
  - Definition of size and location: The objective is to calculate the optimal, or near-optimal lengths of the different buffers to define, i.e., the ideal values that would allow an adequate level of robustness without an excessive capacity consumption. Concurrently, the location in which these buffers are applied can also affect the robustness gain - capacity consumption trade-off. Indeed, it may not be the same to apply a buffer to the beginning of a train's itinerary or at its end. Furthermore, some trains sequences may require more buffer times than others.
  - Application: The way in which the buffers can be considered in a microscopic approach is not unique and may impact the difficulty of the instances. Extensive studies may be required to determine the best option. At least three ways can be thought of: Imposition, the ideal buffer times must be respected. Penalization, the objective value is penalized when the ideal buffers are not observed. A combination of both: A minimum buffer time is defined and always respected, the objective value is penalized if the ideal buffers are not respected.
- *Stochastic optimization.* It consists in the use of stochastic optimization meth-

ods to process timetables capable of absorbing perturbations. These methods are based on the consideration of random variables, which, in this context, emulate the unpredictable emergence of small disturbances during railway operations (e.g., longer dwell times than anticipated). Stochastic optimization may be more suitable than robust optimization for the problems considered in this thesis: It is possible to conjecture that considering the worst case scenario, as the latter does, would imply a too strong inefficiency of the use of capacity. Indeed this conjecture should be carefully verified.

Bibliography

Abbink, E., B. Van den Berg, L. Kroon, and M. Salomon (2004). "Allocation of railway rolling stock for passenger trains". In: *Transportation Science* 38.1, pp. 33–41.

Abbink, E., M. Fischetti, L. Kroon, G. Timmer, and M. Vromans (2005). "Reinventing crew scheduling at Netherlands Railways". In: *Interfaces* 35.5, pp. 393–401.

Abril, M., F. Barber, L. Ingolotti, M. Salido, P. Tormos, and A. Lova (2008). "An assessment of railway capacity". In: *Transportation Research Part E: Logistics and Transportation Review* 44.5, pp. 774 –806.

Albrecht, A., D. Panton, and D. Lee (2013). "Rescheduling rail networks with maintenance disruptions using Problem Space Search". In: *Computers & Operations Research* 40.3, pp. 703 –712.

Arenas, D., R. Chevrier, C. Dhaenens, S. Hanafi, and J. Rodriguez (2013). "Application of a Co-evolutionary Genetic Algorithm to solve the Periodic Railway Timetabling Problem". In: *IESM 2013, 5th international conference on industrial engineering and system management*. Rabat, Morocco, 10p.

Arenas, D., R. Chevirer, S. Hanafi, and J. Rodriguez (2014). "Solving the Periodic Timetabling Problem using a Genetic Algorithm". In: *PANAM 2014, XVIII Congreso Panamericano de Ingeniería de Transito, Transporte y Logistica*. Santander, Spain, 15p.

— (2015). "Solving the Train Timetabling Problem, a mathematical model and a genetic algorithm solution approach". In: *RailTokyo 2015, 6th International Seminar on Railway Operations Modelling and Analysis*. Tokyo, Japan, 12p.

Arenas, D., P. Pellegrini, G. Marlière, and J. Rodriguez (2016a). *RECIFE Class Manager*. Research Report. 19 pages. IFSTTAR.

Arenas, D., P. Pellegrini, S. Hanafi, and J. Rodriguez (2016b). "Timetable Rearrangement to Cope with Railway Maintenance Activities". Under revision, submitted to Computers & Operations Research. 33p.

— (2016c). "Train Timetabling during Infrastructure Maintenance Activities". In: *IEEE International Conference on Intelligent Rail Transportation*, 6p.

Bešinović, N., R. M. Goverde, E. Quaglietta, and R. Roberti (2016). "An integrated micro–macro approach to robust railway timetabling". In: *Transportation Research Part B: Methodological* 87, pp. 14 –32.

Borndörfer, R., T. Schlechte, and E. Swarat (2012). "Railway Track Allocation – Simulation, Aggregation, and Optimization". In: *Proc. 1st International Workshop on High-speed and Intercity Railways (IWHIR 2011)* 2.148, pp. 53 –70.

Brasme, C. (2012). *Travaux Tours Bordeaux, Service Annuel 2012*. Tech. rep. Direction de la Circulation Ferroviaire.

Budai, G., D. Huisman, and R. Dekker (2004). "Scheduling preventive railway maintenance activities". In: *Systems, Man and Cybernetics, 2004 IEEE International Conference on*. Vol. 5, 4171–4176 vol.5.

Bussieck, M. R., P. Kreuzer, and U. T. Zimmermann (1997). "Optimal lines for railway systems". In: *European Journal of Operational Research* 96.1, pp. 54 –63.

Cacchiani, V. and P. Toth (2012). "Nominal and robust train timetabling problems". In: *European Journal of Operational Research* 219.3. Feature Clusters, pp. 727 –737.

Cacchiani, V., A. Caprara, and P. Toth (2008). "A column generation approach to train timetabling on a corridor". English. In: *4OR* 6.2, pp. 125–142.

Cacchiani, V., D. Huisman, M. Kidd, L. Kroon, P. Toth, L. Veelenturf, and J. Wagenaar (2014). "An overview of recovery models and algorithms for real-time

railway rescheduling". In: *Transportation Research Part B: Methodological* 63.0, pp. 15 –37.

Cacchiani, V., A. Caprara, L. Galli, L. Kroon, G. Maróti, and P. Toth (2012). "Railway rolling stock planning: Robustness against large disruptions". In: *Transportation Science* 46.2, pp. 217–232.

Caimi, G., F. Chudak, M. Fuchsberger, M. Laumanns, and R. Zenklusen (2011). "A New Resource-Constrained Multicommodity Flow Model for Conflict-Free Train Routing and Scheduling". In: *Transportation Science* 45.2, pp. 212–227. eprint: http://dx.doi.org/10.1287/trsc.1100.0349.

Caprara, A., M. Fischetti, and P. Toth (2002). "Modeling and Solving the Train Timetabling Problem". In: *Operations Research* 50.5, pp. 851–861.

Caprara, A., L. Galli, and P. Toth (2011). "Solution of the Train Platforming Problem". In: *Transportation Science* 45.2, pp. 246–257. eprint: http://dx.doi.org/10.1287/trsc.1100.0366.

Caprara, A., M. Fischetti, P. Toth, D. Vigo, and P. L. Guida (1997). "Algorithms for railway crew management". In: *Mathematical Programming* 79.1, pp. 125–141.

Caprara, A., M. Monaci, P. Toth, and P. L. Guida (2006). "A Lagrangian heuristic algorithm for a real-world train timetabling problem". In: *Discrete Applied Mathematics* 154.5. {IV} ALIO/EURO Workshop on Applied Combinatorial Optimization, pp. 738 –753.

Chang, Y.-H., C.-H. Yeh, and C.-C. Shen (2000). "A multiobjective model for passenger train services planning: application to Taiwan's high-speed rail line". In: *Transportation Research Part B: Methodological* 34.2, pp. 91 –106.

Charnes, A. and M. H. Miller (1956). "A Model for the Optimal Programming of Railway Freight Train Movements". In: *Management Science* 3.1, pp. 74–92. eprint: http://dx.doi.org/10.1287/mnsc.3.1.74.

CIT (2014). *European General Terms and Conditions of use of railway infrastructure*. International Rail Transport Committee.

Claessens, M., N. van Dijk, and P. Zwaneveld (1998). "Cost optimal allocation of rail passenger lines". In: *European Journal of Operational Research* 110.3, pp. 474 –489.

Cordeau, J.-F., P. Toth, and D. Vigo (1998). "A Survey of Optimization Models for Train Routing and Scheduling". In: *Transportation Science* 32.4, pp. 380–404. eprint: http://dx.doi.org/10.1287/trsc.32.4.380.

Corman, F., A. D'Ariano, I. A. Hansen, and D. Pacciarelli (2011). "Optimal multi-class rescheduling of railway traffic". In: *Journal of Rail Transport Planning & Management* 1.1. Robust Modelling of Capacity, Delays and Rescheduling in Regional Networks, pp. 14 –24.

Şahin, G. and B. Yüceoğlu (2011). "Tactical crew planning in railways". In: *Transportation Research Part E: Logistics and Transportation Review* 47.6, pp. 1221 –1243.

D'Ariano, A., D. Pacciarelli, and M. Pranzo (2007). "A branch and bound algorithm for scheduling trains in a railway network". In: *European Journal of Operational Research* 183.2, pp. 643 –657.

Dollevoet, T., D. Huisman, M. Schmidt, and A. Schöbel (2012). "Delay Management with Rerouting of Passengers". In: *Transportation Science* 46.1, pp. 74–89. eprint: http://dx.doi.org/10.1287/trsc.1110.0375.

Edwards, A. L. (1976). "The Correlation Coefficient." In: *An Introduction to Linear Regression and Correlation*. W. H. Freeman. Chap. 4, pp. 33–46.

Fischer, F., C. Helmberg, J. Janßen, and B. Krostitz (2008). "Towards Solving Very Large Scale Train Timetabling Problems by Lagrangian Relaxation". In: *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*. Ed. by M. Fischetti and P. Widmayer. Vol. 9. OpenAccess Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Forsgren, M., M. Aronsson, and S. Gestrelius (2013). "Maintaining tracks and traffic flow at the same time". In: *Journal of Rail Transport Planning & Management* 3.3, pp. 111 –123.

Fu, H., L. Nie, L. Meng, B. R. Sperry, and Z. He (2015). "A hierarchical line planning approach for a large-scale high speed rail network: The China case". In: *Transportation Research Part A: Policy and Practice* 75, pp. 61 –83.

Fukumori, K. (1980). *Fundamental Scheme for Train Scheduling (Application of Range-Constriction Search).* Tech. rep. DTIC Document.

Fukumori, K., H. Sano, T. Hasegawa, and T. Sakai (1987). "Fundamental algorithm for train scheduling based on artificial intelligence". In: *Systems and Computers in Japan* 18.3, pp. 52–64.

García-Archilla, B., A. J. Lozano, J. A. Mesa, and F. Perea (2013). "GRASP algorithms for the robust railway network design problem". In: *Journal of Heuristics* 19.2, pp. 399–422.

Gattermann, P., P. Großmann, K. Nachtigall, and A. Schöbel (2016). "Integrating Passengers' Routes in Periodic Timetabling: A SAT approach". In: *16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016).* Ed. by M. Goerigk and R. Werneck. Vol. 54. OpenAccess Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 1–15.

Gomes, C. P., H. Kautz, A. Sabharwal, and B. Selman (2008). "Satisfiability solvers". In: *Foundations of Artificial Intelligence* 3, pp. 89–134.

Goolen, J. (2008). "La Massification de la Maintenance, Une véritable démarche sociotechnique". In: *Revue Générale des Chemins de Fer* 5, pp. 17–33.

Goossens, J.-W., S. van Hoesel, and L. Kroon (2006). "On solving multi-type railway line planning problems". In: *European Journal of Operational Research* 168.2. Feature Cluster on Mathematical Finance and Risk Management, pp. 403 –424.

Hachemane, P. (1997). "Évaluation de la capacité de réseaux ferroviaires." PhD thesis. École Polytechnique Fédérale de Lausanne.

Hansen, I. and J. Pachl (2014). *Railway Timetabling & Operations: Analysis - Modelling - Optimisation - Simulation - Performance Evaluation.* Eurailpress.

Higgins, A (1998). "Scheduling of railway track maintenance activities and crews". In: *Operational Research Society* 49, pp. 1026–1033.

Higgins, A., E. Kozan, and L. Ferreira (1996). "Optimal scheduling of trains on a single line track". In: *Transportation Research Part B: Methodological* 30.2, pp. 147 –161.

— (1997). "Heuristic Techniques for Single Line Train Scheduling". English. In: *Journal of Heuristics* 3.1, pp. 43–62.

Hooghiemstra, J. S., L. G. Kroon, M. A. Odijk, M. Salomon, and P. J. Zwaneveld (1999). "Decision Support Systems Support the Search for Win-Win Solutions in Railway Network Design". In: *Interfaces* 29.2, pp. 15–32. eprint: http://pubsonline.informs.org/doi/pdf/10.1287/inte.29.2.15.

IBM (2015). *IBM ILOG CPLEX Optimization Studio CPLEX User Manual Version 12 release 6*. IBM Corporation.

Ingolotti, L., A. Lova, F. Barber, P. Tormos, M. Salido, and M. Abril (2006). "New Heuristics to Solve the "CSOP" Railway Timetabling Problem". English. In: *Advances in Applied Artificial Intelligence*. Ed. by M. Ali and R. Dapoigny. Vol. 4031. Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 400–409.

Jespersen-Groth, J., D. Potthoff, J. Clausen, D. Huisman, L. Kroon, G. Maróti, and M. N. Nielsen (2009). "Disruption Management in Passenger Railway Transportation". In: *Robust and Online Large-Scale Optimization: Models and Techniques for Transportation Systems*. Ed. by R. K. Ahuja, R. H. Möhring, and C. D. Zaroliagis. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 399–421.

Jütte, S. and U. W. Thonemann (2012). "Divide-and-price: A decomposition algorithm for solving large railway crew scheduling problems". In: *European Journal of Operational Research* 219.2, pp. 214 –223.

Klabes, S. G. (2010). "Algorithmic railway capacity allocation in a competitive European railway market". Aachen, Techn. Hochsch., Diss., 2010. PhD thesis. Aachen, 209 S. : graph. Darst.

Kraft, E. R. (1982). "Jam Capacity of Single Track Rail". In: *Proceedings of the Transportation Research Forum*. Vol. 23. Transportation Research Forum 1, pp. 461–471.

Kroon, L., D. Huisman, E. Abbink, P.-J. Fioole, M. Fischetti, G. Maróti, A. Schrijver, A. Steenbeek, and R. Ybema (2009). "The New Dutch Timetable: The OR Revolution". In: *Interfaces* 39.1, pp. 6–17. eprint: http://interfaces.journal.informs.org/content/39/1/6.full.pdf+html.

Kroon, L., D. Huisman, and G. Maroti (2007). *Railway timetabling from an operations research Perspective*. Econometric Institute Report EI 2007-22. Erasmus University Rotterdam, Econometric Institute.

Krueger, H. (1999). "Parametric Modeling in Rail Capacity Planning". In: *Proceedings of the 31st Conference on Winter Simulation: Simulation—a Bridge to the Future - Volume 2*. WSC '99. Phoenix, Arizona, USA: ACM, pp. 1194–1200.

Kümmling, M, P Großmann, K Nachtigall, J Opitz, and R Weiß (2013). "A State-of-the-art Realization of Automatic Railway Timetable Computation". In: *Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transport Systems (MT-ITS 2013), December*, pp. 2–4.

Kwan, R. and P. Mistry (2003). *A Co-evolutionary Algorithm for Train Timetabling*. Tech. rep. University of Leeds.

Lafont, J. and M. Antoni (2012). "Optimisation analytique du processus "IPCS"". In: *18 me Congrès de Maîtrise des Risques et Sûreté de Fonctionnement*.

Liebchen, C. (2008). "The First Optimized Railway Timetable in Practice". In: *Transportation Science* 42.4, pp. 420–435.

Liebchen, C., M. Schachtebeck, A. Schöbel, S. Stiller, and A. Prigge (2010). "Computing delay resistant railway timetables". In: *Computers & Operations Research* 37.5. Disruption Management, pp. 857 –868.

Lusby, R. M., J. Larsen, M. Ehrgott, and D. Ryan (2011). "Railway track allocation: models and methods". English. In: *OR Spectrum* 33.4, pp. 843–883.

Magnanti, T. L. and R. T. Wong (1984). "Network Design and Transportation Planning: Models and Algorithms". In: *Transportation Science* 18.1, pp. 1–55. eprint: http://dx.doi.org/10.1287/trsc.18.1.1.

Maróti, G. (2006). "Operations Research Models for Railway Rolling Stock Planning". PhD thesis. Eindhoven University of Technology, Eindhoven, The Netherlands.

Nachtigall, K. and S. Voget (1996). "A genetic algorithm approach to periodic railway synchronization". In: *Computers & Operations Research* 23.5, pp. 453 –463.

Odijk, M. A. (1996). "A constraint generation algorithm for the construction of periodic railway timetables". In: *Transportation Research Part B: Methodological* 30.6, pp. 455 –464.

Oliveira, E. and B. M. Smith (2000). *A Job-Shop Scheduling Model for the Single-Track Railway Scheduling Problem*. Tech. rep. 2000.21. HTTP://WWW.OPENTRACK.CH: University of Leeds.

Pellegrini, P., G. Marliere, and J. Rodriguez (2014). "Optimal train routing and scheduling for managing traffic perturbations in complex junctions". In: *Transportation Research Part B: Methodological* 59.0, pp. 58 –80.

Pellegrini, P., G. Marlière, R. Pesenti, and J. Rodriguez (2015). "RECIFE-MILP: An Effective MILP-Based Heuristic for the Real-Time Railway Traffic Management Problem". In: *IEEE Transactions on Intelligent Transportation Systems* 16.5, pp. 2609–2619.

Peng, F., S. Kang, X. Li, Y. Ouyang, K. Somani, and D. Acharya (2011). "A Heuristic Approach to the Railroad Track Maintenance Scheduling Problem". In: *Computer-Aided Civil and Infrastructure Engineering* 26.2, pp. 129–145.

Profillidis, V. A. (2014). *Railway management and engineering*. Ashgate Publishing, Ltd.

Putallaz, Y. (2007). "Gestion Strategique de la maintenance et de la capacite d'un reseau ferre". PhD thesis. Ecole Polytechnique Federale de Lausanne.

Putallaz, Y. and R. Rivier (2003). "Strategic maintenance and renewal policy of a railway corridors, taking into account the value of capacity". In: *World Conference on Rail Research, WCRR 03*. Edinburgh.

Putallaz, Y. and P. Tzieropoulos (2012). *Audit sur l'état du réseau (Audit Rivier)*. Tech. rep. Ecole Polytechnique Federale de Lausanne.

Retiveau, R. (1987). *La Signatiation Ferroviaire*. Ed. by P. de l'École Nationale des Ponts et Chaussées. Presses de l'École Nationale des Ponts et Chaussées.

RFF (2009). *Referentiel de Repartition de la Capacite d'Infrastructure*. Reseau Ferre de France.

— (2011). *Referentiel Capacite - Description des principes et processus de la gestion de la capacite d'Infrastructure*. Reseau Ferre de France.

Rivier, R. and Y. Putallaz (2005). *Audit sur létat du Réseau Ferré National Francais*. Tech. rep. Ecole Polytechnique Federale de Lausanne.

Rodriguez, J., X. Delorme, X. Gandibleux, G. Marliere, R. Bartusiak, F. Degoutin, and S. Sobieraj (2007). "RECIFE: Models and tools for analyzing rail capacity". In: *RTS-Recherche Transports Securite* 95, pp. 129–146.

Samà, M., P. Pellegrini, A. D'Ariano, J. Rodriguez, and D. Pacciarelli (2016). "Ant colony optimization for the real-time train routing selection problem". In: *Transportation Research Part B: Methodological* 85, pp. 89 –108.

Schlechte, T. (2012). *Railway Track Allocation: Models and Algorithms*. Südwestdeutscher Verlag für Hochschulschriften AG Company KG.

Schlechte, T., R. Borndörfer, B. Erol, T. Graffagnino, and E. Swarat (2011). "Micro–macro transformation of railway networks". In: *Journal of Rail Transport Planning & Management* 1.1. Robust Modelling of Capacity, Delays and Rescheduling in Regional Networks, pp. 38 –48.

Schöbel, A. (2012). "Line planning in public transportation: models and methods". In: *OR Spectrum* 34.3, pp. 491–510.

Serafini, P. and W. Ukovich (1989). "A mathematical model for periodic scheduling problems". In: *SIAM J. Discret. Math.* 2.4, pp. 550–581.

Siebert, M. and M. Goerigk (2013). "An experimental comparison of periodic timetabling models". In: *Computers & Operations Research* 40.10, pp. 2251 –2259.

SNCF (1991). *Installations de Contre-Sens*. Societe National des Chemins de Fer.

Sourd, F. (2010). "SIOUCS: un outil d'aide à la décision pour la gestion des circulations". In: *Revue Générale des Chemins de Fer* 12, pp. 20–26.

Steinzen, I., V. Gintner, L. Suhl, and N. Kliewer (2010). "A time-space network approach for the integrated vehicle-and crew-scheduling problem with multiple depots". In: *Transportation Science* 44.3, pp. 367–382.

Szpigel, B. (1972). "Optimal train scheduling on a single track railway". In: *Operational Research '72*, pp. 343–351.

Theeg, G., E. Anders, and S. Vlasenko (2009). *Railway Signalling & Interlocking: International Compendium*. Eurailpress.

Tormos, P., A. Lova, L. Ingolotti, and F. Barber (2008). *A Genetic approach to Robust Train Timetabling*. Tech. rep. ARRIVAL Project.

Veelenturf, L., L. Nielsen, G Maróti, and L. Kroon (2011). "Passenger oriented disruption management by adapting stopping patterns and rolling stock schedules". In: *the 4th International Seminar on Railway Operations Modelling and Analysis-RailRome*.

Wong, R. C. W., T. W. Y. Yuen, K. W. Fung, and J. M. Y. Leung (2008). "Optimizing Timetable Synchronization for Rail Mass Transit". In: *Transportation Science* 42.1, pp. 57–69.

Zhou, X. and M. Zhong (2007). "Single-track train timetabling with guaranteed optimality: Branch-and-bound algorithms with enhanced lower bounds". In: *Transportation Research Part B: Methodological* 41.3, pp. 320 –341.

Acronyms

CP Constraint Programming. 55, 56, 61, *Glossary:* constraint programming

FCFS First-Come First-Served. 104–106, *Glossary:* first-come first-served

GA Genetic Algorithm. 56, 58, 60, *Glossary:* genetic algorithm

IM Infrastructure Manager. 13–17, 19, 20, 23, 25, 31, 44, 50, 56, 83, 88, 94, 141, 143, *Glossary:* infrastructure manager

IPCS Installations Permanentes de Contre-Sens. 30, 32, 105–107, *Glossary:* Installations Permanentes de Contre-Sens

MA Maintenance Activity. 16, 19, 22, 27–33, 97–105, 107–111, 116, 117, 119, 122–126, 128–134, 138–145, *Glossary:* maintenance activity

MILP Mixed-Integer Linear Programming. 5, 53, 58, 62, 65, 70, 81, 82, 84, 94, 98, 107, 119, 123, 129, *Glossary:* mixed-integer linear programming

MT Maintenance Train. 97, 98, 101–103, 107, 109, 110, 116, 123, 124, 136, 138, 139, 142–144, *Glossary:* maintenance train

OT Operational Train. 66, 70, 71, 76, 102–104, 112, 114, 115, 123, *Glossary:* operational train

PESP Periodic Event Scheduling Problem. 59–62, *Glossary:* periodic event scheduling problem

RFF Réseau Ferré de France (French Railway Network). 20

rtRTMP real-time Railway Traffic Management Problem. 19, 69, *Glossary:* real-time Railway Traffic Management Problem

RU Railway Undertaking. 13–17, 19–23, 25, 26, 49, 50, 56, 65, 67, 102, 141, 143, *Glossary:* railway undertaking

SNCF Societé Nationale des Chemins de fer Francais (National Society of French Railways). 20, 25, 26, 29, 83, 88

TDS Track Detection Section. 38–45, 65, 70–73, 75–81, 83, 102, 108, 110–119, 122, *Glossary:* track detection section

TSL Temporary Speed Limitation. 102, 103, 107, 108, 111, 113, 117, 124, 126, 136, 138, 139, 142–145, *Glossary:* temporary speed limitation

TTP Train Timetabling Problem. 17, 24, 25, 45–47, 49, 51–55, 57, 59, 61, 62, 64, 65, 93, 94, 99, 142, 144, *Glossary:* train timetabling problem

Glossary

actual running time It denotes the running time that is actually considered for a train that may be affected by a TSL imposed by a MA. 111, 113–115

annual railway service plan A compound of documents that specify several aspects of the railway service for a year of operations, e.g., the annual commercial timetable, the capacity allocation plan, the crew schedules, etc. 12, 14–16, 18–22, 25, 33

block section A section of track which can be utilized by only one train at any given time. Its length is variable and it is composed by one or more TDSs. 38–40, 43, 45, 63, 65, 71, 78–80, 83

buffer A time supplement to the running, dwelling, turnaround, etc., that is mainly defined to increase the robustness of a timetable. 50, 67, 91–95, 144

capacity The maximum number of trains that would be able to operate on a given railway infrastructure, during a specific time interval, given a set of operational conditions and a given service quality. 11, 15, 17, 19, 22–27, 29, 30, 32–34, 36, 37, 39, 41, 46, 49, 52, 63, 65, 67, 68, 94, 95, 97, 98, 102, 104, 105, 138, 139, 141–143

clearing time The time elapsed between the exit of the head of a train from a TDS, and the exit of its rear. 42, 72, 79, 109–111, 115

commercial timetable A type of timetable mainly defined for the users of the railway system, e.g., passengers, freight companies, etc. It defines the scheduled arrival and departure times of trains at stations. 23

conflict In timetabling, a conflict exists when two trains travelling at the planned speed would concurrently require the same track segment. 24, 36, 43, 49, 97, 105, 122, 123, 137–139, 144

constraint programming A programming paradigm wherein relations between variables are stated in the form of constraints. 55

control point The locations or relevant points, such as stations, where a train departs from, arrives to, and perform intermediate stops, if any. 41, 66, 70–72, 74–77, 83, 93, 94, 110, 114, 115, 144

control zone The portion of the railway infrastructure that is considered for timetabling procedures. The limits of the control zone do not necessarily begin and end at stations. 41, 43, 65, 68, 69, 74, 75, 77, 83, 88, 91–94, 124, 132, 143, 144

default route The route assigned to a train in a timetable. 102, 109, 119, 125, 126, 139, 145

directly affected TDS The set of consecutive TDSs where a MA takes place. 108, 110, 116, 119

dwell time The time that a train remains in a platform to allow passengers to disembark, transfer and board the train. 41, 42, 62, 66–68, 71, 73, 76, 90–95, 112, 143, 144

feasible A timetable is considered feasible if all planned train circulations are conflict-free and respect the capacity and safety constraints of the system. 24, 46, 49, 62, 64, 65, 94, 97, 103

first-come first-served In a railway scheduling context, is a scheduling policy in which trains are scheduled in the order they arrived/entered to the considered control zone, without other biases or preferences. 104

formation time The necessary time to set and lock all TDSs inside a block section in the desired position, plus an additional time to take into account the signal visibility distance. 43, 78

GAP Is the percentage difference between the best feasible solution found by an algorithm within a run and the best bound identified. 86–88, 90, 130–132, 134, 135, 139, 145

genetic algorithm Is a search meta-heuristic that mimics the process of natural selection. Is generally used to generate useful solutions to optimization and search problems. 56

green wave policy A train circulating policy which implies that a train must follow its itinerary without having to stop or reduce its speed due to a conflict with another train. 24, 26, 66, 69, 71, 102, 103, 114, 123

headway time The minimum time interval that must separate the entrance in a track of two trains using it. It can also be applied to the exit of the track. 36, 104

horizon The period of time considered in an instance of a (re)timetabling problem. It determines the trains to be considered in the instance: Those trains that enter the control zone between the start and end times of the horizon. 83, 125, 132, 133

inbound MT A MT that must arrive to the MA location at the exact time as the MA begins. 109

infrastructure manager any body or firm responsible for establishing, managing and maintaining railway infrastructure, including traffic management, control-command and signalling. 13

Installations Permanentes de Contre-Sens Permanent opposite direction circulation equipments, allows the circulation of trains in both directions over a track segment. 30

itinerary It defines a train's origin, destination and, in some cases, a set of intermediate stops. 15, 16, 23, 24, 41, 62, 63, 66, 67, 74, 93

maintenance activity The set of actions performed for retaining the railway infrastructure in, or restoring it to, a state in which it can perform its required function. It includes the construction of new infrastructure. 16

maintenance train A particular train whose rolling stock is specifically equipped to perform MAs or to transport maintenance materials. 97

mixed-integer linear programming Is a very general framework for capturing problems with both discrete decisions and continuous variables. 5

occupation time The time elapsed between the entrance of the head of a train into a TDS and the exit of its rear. 41, 42, 75, 112, 114

operational train A train that is planned during the timetabling procedure, in the tactical level, and must respect the green wave policy. 66

outbound MT A MT that must depart from the MA location as soon as the MA is finished. 109

periodic event scheduling problem A generic model of a scheduling problem that is used for the scheduling of periodic activities. 59

railway system The conjunction of all elements that allow rail transportation, i.e., infrastructure, rolling stock, crews, operational conditions, schedules, etc. 9–13, 23, 24, 26, 27, 29, 34, 37, 44, 46, 49, 52, 82, 141, 142

railway undertaking any public or private entity that provide services for the transport of goods and/or passengers by rail. 13

real-time Railway Traffic Management Problem The problem of rescheduling and re-routing trains in real-time to cope with operational perturbations.. 19

release time The necessary time to unlock and reset all TDSs inside a block section to their default position. 43, 79

reservation time The time elapsed between the start of the reservation and the actual occupation of a TDS by a train. 43

route A complete sequence of TDSs which can be traversed by a train to fulfil its itinerary. 41, 42, 62, 65, 66, 70–73, 75, 76, 78–81, 83, 85, 87–90, 94, 102, 104, 110–114, 119, 120, 122–124, 144

running time The time necessary for a train to traverse a segment of the railway infrastructure. This segment may be the tracks connecting two stations or a single TDS, in macroscopic and microscopic approaches, respectively. 36, 42, 66, 72, 75, 77, 79, 102, 108–111, 113, 114, 138, 144

running time mode The mode of running time considered for a train traversing a TDS, it my be restricted or nominal, depending if a TSL is imposed or not. 111, 114, 115, 117

signal A semaphore governing the entrance to a block section, it gives information to the drivers about the utilization of the forthcoming block sections. 38, 39, 65, 103

temporary speed limitation A safety measure during the performance of some MAs. It consists in reducing the allowed maximum speed of trains circulating on tracks adjacent to the MA locations. 102

track detection section A section of a track where the presence of a train can be automatically detected. 38

train battery A group of trains circulating in the same direction over the same tracks. 104

train timetabling problem The problem of producing a train timetable that meets a set of commercial needs, safety and capacity constraints. 17

train-path request A train service required by a RU. It defines the itinerary and desired schedule of the required train. 23, 25, 26, 62–68, 71, 73–76, 83, 88, 91–94, 107, 108, 143, 144

travel time The time elapsed between the entrance and the exit of a train in the control zone, i.e., the time needed for a train to complete its itinerary. 41, 67, 90

TSL affected TDS The set of TDSs, typically contiguous to the directly affected TDSs, where train circulations are restricted to a TSL. 108–110, 116

utilization time The period of time in which a train utilizes a TDS, it is defined by the sum of the formation, reservation, occupation and release times of the concerned TDS. 43, 79, 111, 112, 115

working timetable A type of timetable defined for the railway's operating staff, i.e., IMs and RUs. It includes timing details at every station, major junction or other significant locations in the railway network. 23, 24