



HAL
open science

Construction of a methodology and associated tool for the design and planning of a bus network

Hector Gatt

► **To cite this version:**

Hector Gatt. Construction of a methodology and associated tool for the design and planning of a bus network. Operations Research [math.OC]. Ecole nationale supérieure Mines-Télécom Atlantique, 2023. English. NNT: 2023IMTA0355 . tel-04532087

HAL Id: tel-04532087

<https://theses.hal.science/tel-04532087>

Submitted on 4 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE
MINES-TÉLÉCOM ATLANTIQUE BRETAGNE
PAYS-DE-LA-LOIRE - IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 648
Sciences pour l'Ingénieur et le Numérique
Spécialité : *Informatique*

Par

Hector GATT

**Construction d'une méthodologie et d'un outil associé pour la
conception et la planification d'un réseau de bus**

Thèse présentée et soutenue à IMT Atlantique Nantes, le 21 mars 2023
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N)
Thèse N° : 2023IMTA0355

Rapporteurs avant soutenance :

Luce BROTCORNE Directrice de recherche à INRIA
Dominique FEILLET Professeur aux Mines Saint-Etienne

Composition du Jury :

Président :	Laurent ALFANDARI	Professeur à ESSEC Business School
Examineurs :	Luce BROTCORNE	Directrice de recherche à INRIA
	Dominique FEILLET	Professeur aux Mines Saint-Etienne
	Aurélien FROGER	Maître de conférences à Université de Bordeaux
Dir. de thèse :	Fabien LEHUÉDÉ	Professeur à IMT Atlantique
Co-encadrant de thèse :	Arnaud LAURENT	Maître de conférences à Nantes Université
Co-encadrant de thèse :	Thomas G. YEUNG	Ingénieur de recherche à Amazon

Invité(s) :

Jean-Marie FRECHE Directeur Technique à Lumiplan

Lumiplan® 

RÉSUMÉ EN FRANÇAIS

Le transport public est un maillon indispensable de notre société. Celui-ci prodigue en effet à la fois une mobilité écologique et sociale. Néanmoins, le réseau de transport public se doit d'être à la fois maîtrisé en terme de coûts d'exploitation et de qualité pour les passagers. Pour atteindre cet objectif, les outils de la recherche opérationnelle peuvent aider à de nombreux niveaux de décision dans la conception, la planification et la gestion opérationnelle d'un réseau. Le travail présenté dans cette thèse s'inscrit dans cette démarche de conception et de planification en amont des réseaux de transport public. Ce travail a été réalisé dans le cadre d'une Convention Industrielle de Formation par la REcherche (CIFRE) impliquant le laboratoire de recherche LS2N et la société industrielle Lumiplan.

Selon Ceder and Wilson [1986], la planification d'un réseau de transport se fait en cinq étapes. Ces étapes vont des décisions stratégiques à long terme aux décisions opérationnelles à court terme. Au niveau stratégique, la première étape concerne la conception du réseau. Cette étape comporte elle même plusieurs aspects : création d'arrêts, création de nouveaux tronçons, détermination de lignes sur l'ensemble des arrêts et tronçons utilisables. Dans le cadre de nos travaux nous nous concentrons sur la détermination des lignes. À mi-chemin entre le niveau stratégique et le niveau tactique, se positionne l'étape de la détermination des fréquences. Cette étape intervient après la conception du réseau et à intervalles réguliers afin de s'adapter aux variations dans la demande des voyageurs. La détermination des horaires et la création des plannings matériels et conducteurs, communément appelées étapes de graphicage et d'habillage, sont toutes trois des étapes de niveau opérationnel. Celles-ci sont effectuées de manière saisonnière pour adapter les plannings aux fluctuations de demandes. L'entreprise Lumiplan propose actuellement le logiciel Heurès aux entreprises de transport public pour soutenir leurs opérations de planification des services matériels et conducteurs. Basé sur un partenariat entre IMT Atlantique et la société Lumiplan, ces travaux s'inscrivent dans une volonté de l'entreprise Lumiplan de proposer, en complément du logiciel Heurès, un outil d'optimisation des réseaux de bus au niveau des étapes de conception de lignes et de détermination des fréquences.

La satisfaction de la demande des passagers avec une bonne qualité de service et à coût réduit dépend à la fois du tracé des lignes et de la fréquence sur ces lignes aux différentes périodes d'exploitation de la journée. La problématique générale est donc un problème intégré de conception de lignes et de détermination des fréquences. Néanmoins, la trop grande complexité du problème nous pousse à séparer le problème en deux problèmes distincts. Ces deux problèmes sont connus dans la littérature sous les noms de Line Planning Problem (LPP) et Frequency Setting Problem (FSP). Dans cette thèse ils sont définis de la manière suivante :

- Le LPP cherche à déterminer un ensemble de lignes à opérer en considérant une fréquence globale sur la journée d'exploitation pour chacune de ces lignes. Les itinéraires des passagers sont modélisés comme des flux s'écoulant sans stratégie individuelle sur le réseau.
- Le FSP cherche à affiner les fréquences d'exploitation sur plusieurs périodes de la journée d'exploitation en intégrant les stratégies individuelles des passagers. Pour cela on s'intéressera à la modélisation des temps d'attente.

Pour résoudre ces problèmes, nous définissons des approches heuristiques basées sur la programmation mathématique. L'approche adoptée dans cette thèse se différencie d'un grand nombre d'approches de la littérature dans le sens où nous considérons d'une part un objectif économique, inhérent à la vision *opérateur* du réseau, combiné à la définition de contraintes sur le niveaux de service offert aux passagers.

Le processus de résolution du line planning problem combine des méthodes de génération de colonnes et d'énumération de colonnes. Ces deux méthodes permettent de générer simultanément lignes de bus et chemins de passagers. La plupart des approches de la littérature sont basées sur la sélection de ligne dans un ensemble généré en amont. Ces approches sont limitées par la taille de l'ensemble de lignes qu'il est possible d'intégrer à un algorithme d'optimisation. Les approches considèrent donc un ensemble de lignes de taille restreinte, influant dès lors sur le réseau obtenu en sortie. Les expérimentations montrent que les avancées réalisées dans cette thèse, sur la création de lignes par génération de colonnes, permet de concevoir de meilleurs réseaux.

Le processus de détermination des fréquences se base sur la sélection de chemins de passagers à partir d'un ensemble préalablement généré. Un enjeu majeur du problème est la modélisation du comportement des passagers. L'hypothèse retenue dans ce travail est qu'un passager emprunte un trajet de durée minimum sur le réseau. Plus précisément, pour chaque paire OD, les trajets possibles sont modélisés comme des chemins. Chaque

chemin est défini par une séquence d'arcs, chacun associé à une ligne. La durée d'un chemin est définie comme la somme du temps de trajet sur ce chemin et du temps d'attente estimé induit par la montée au premier arrêt et lors des transferts. Ce temps d'attente estimé dépend des fréquences déployées sur les lignes. Le problème de détermination des fréquences déployées est donc un problème bi-niveaux où le leader est l'exploitant du réseau de bus qui fixe les fréquences des lignes en minimisant ses coûts et les followers sont les passagers qui minimisent la durée de leurs trajets en fonction des fréquences des lignes. Le problème bi-niveaux est reformulé sous la forme d'un problème mono-niveau comportant un nombre quadratique de variables et contraintes dépendant du nombre de chemins possibles sur le réseau.

Nos contributions pour la partie scientifique pour le LPP sont les suivantes :

- Nous étendons un modèle de la littérature [Borndörfer and Karbstein, 2012] pour résoudre le LPP et le renforçons avec des contraintes opérationnelles.
- Nous proposons de nouveaux paramètres de niveaux de service pour les passagers et les intégrons dans le modèle renforcé du LPP.
- Nous proposons un processus complet, appelé CGEP (Column Generation and Enumeration Process), combinant les méthodes de génération et d'énumération de colonnes, pour résoudre le LPP avec niveau de service.
- La génération des lignes de bus est directement intégrée dans CGEP et évite ainsi de devoir générer des ensembles de lignes à l'aide de méthodes tierces en amont de notre programme comme c'est le cas avec la plupart des méthodologies existantes.

Concernant le FSP, nos contributions pour la partie scientifique sont les suivantes :

- Nous proposons un programme linéaire mixte en nombres entiers pour résoudre le FSP avec des contraintes opérationnelles et de niveaux de service.
- Dans notre modèle de FSP, nous utilisons une contrainte de plus court chemin pour modéliser les déplacements des passagers.
- Pour rendre ce modèle traitable sur des instances réelles, nous proposons un processus de sélection dynamique de chemins de passagers appelé PPSP (Passenger Path Selection Process). Ce processus sélectionne les chemins qui sont intégrés au modèle et permet ainsi une grande différence de temps de résolution comparé à la résolution du modèle avec tous les chemins possibles.

Les méthodes proposées sont évaluées sur un cas d'étude construit à partir de données fournies par la Régie des Transports Poitevin (RTP). Pour cela, à partir d'un réseau et

d'un ensemble de matrices origine-destination tous deux générés à partir des données RTP, nous mettons alors en application nos deux algorithmes CGEP et PPSP. Après avoir introduit un ensemble de règles de filtrage concernant la typologie des lignes, nous observons le potentiel d'amélioration en termes de tracé de lignes que peut fournir l'algorithme CGEP. À partir de ces résultats, nous sommes également en mesure de montrer l'impact que peuvent avoir les paramètres de niveau de service ainsi que de la charge moyenne des bus sur les coûts opérationnels du réseau associé. De plus, nous appliquons également l'algorithme PPSP sur de petites périodes horaires afin d'analyser les changements proposés en termes de fréquences de lignes par rapport aux fréquences fournies par l'exploitant. Les résultats obtenus démontrent la pertinence de nos modèles et algorithmes pour les entreprises de transport public, que ceux-ci soient utilisés pour affiner les fréquences sur un réseau ou pour produire des estimations de coûts et de flotte dans une perspective de ré-optimisation de réseau.

Ce manuscrit de thèse se compose des cinq chapitres suivants :

- Le **Chapitre 1** présente le processus global de planification des réseaux de bus, et plus particulièrement les deux étapes de conception du plan de lignes et de détermination des fréquences des lignes. Dans la littérature, les deux problèmes associés à cette étape sont le Line Planning Problem et le Frequency Setting Problem. Une revue de la littérature traitant de ces deux problèmes est donc faite. Enfin, nous présentons les paramètres de niveaux de service généralement utilisés, qu'ils soient issus de la littérature scientifique ou d'un décret gouvernemental.
- Le **Chapitre 2** présente la méthodologie choisie pour résoudre les deux étapes de la conception du plan de lignes et de la détermination des fréquences des lignes.
- Le **Chapitre 3** est consacré au problème de la conception du plan de lignes. Il présente l'adaptation d'un modèle de LPP issu de la littérature. Les évolutions apportées intègrent l'objectif économique, la spécification de contraintes opérationnelles et deux paramètres de niveau de service. Un processus de résolution basé sur une méthode de génération de colonnes est proposé. Les expérimentations montrent notamment la supériorité de l'approche par rapport aux travaux ayant inspiré le modèle.
- Le **Chapitre 4** est consacré au problème de détermination des fréquences. Nous y proposons un modèle linéaire en nombre entier mono-niveau ayant pour objectif de minimiser les coûts de l'opérateur tout en respectant des contraintes de plus courts chemins en terme de temps de trajet pour les voyageurs. Un processus de résolution

basé sur une sélection dynamique des chemins des voyageurs est également proposé. Enfin, quelques résultats issus d'exemples de la littérature sont décrits.

- Le **chapitre 5** est consacrée au cas d'étude basé sur le réseau de l'agglomération de Poitiers. Ce chapitre met en évidence l'intérêt de notre travail dans le cadre d'une réoptimisation d'un réseau de transport public local.

Le chapitre 3 fait l'objet d'un article soumis pour publication dans une revue internationale. Les résultats de cette recherche ont également été présentés lors des conférences [Gatt et al., 2021] et [Gatt et al., 2022b]. Quant au chapitre 4, des résultats ont été présentés lors des conférence [Gatt et al., 2022a] et [Gatt et al., 2023].

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to everyone who has supported me throughout my PhD years.

First and foremost, I extend my sincerest thanks to my thesis rapporteurs, Luce Brotcorne and Dominique Feillet, for their willingness to evaluate my work, as well as my thesis jury Laurent Alfandari and Aurélien Froger.

I am also immensely grateful to my supervisors, Fabien Lehuédé, Arnaud Laurent, and Thomas Yeung, for their invaluable guidance, continuous support, and patience. Their wealth of knowledge and experience has been instrumental in my academic research and daily life.

I would like to express my appreciation to the entire Lumiplan team, especially Jean-Marie Freche for his trust, enthusiasm, and support for my work.

Additionally, I thank all members of the Modelis team for their kindness and assistance throughout my PhD study.

I am deeply grateful to Marika Karbstein for her generosity in sharing her datasets, which were essential to my research.

Finally, I would like to express my sincere appreciation to the Régie des Transports Poitevins for generously sharing their data, enabling me to produce a comprehensive case study for my thesis.

Last but not least, I would like to express my deepest gratitude to my wife Marie, whose unwavering support, encouragement, and patience have been the driving force behind my academic achievements.

I would also like to extend my thanks to my parents, sister, aunt, grandfather, and my two best friends, Dina and Nicolas, for their constant support and understanding.

TABLE OF CONTENTS

Introduction	17
1 History of urban transportation networks	17
2 Context of the thesis	20
3 Objective of the thesis	23
4 Organization of the thesis	25
1 Literature Review for the Planning of Urban Transportation Networks	27
1 The global planning process	28
2 The Line Planning Problem	29
2.1 Definition	29
2.2 Solving approaches	32
2.2.1 A priori line pool generation approaches	32
2.2.2 Integrated line generation approaches	35
2.3 Service levels	37
2.3.1 Travel time related service levels	38
2.3.2 Transfers related service levels	39
3 The Frequency Setting Problem	40
3.1 Definition	40
3.2 Methods	41
4 Research gaps	44
4.1 Line Planning Problem	45
4.2 Frequency Setting Problem	45
2 Methodology and contributions	47
1 Decomposition choices	47
2 Summary of contributions	50
3 The Line Planning Problem with service level (LPP-SL)	53
1 The Line Planning Problem with service level	54
1.1 Problem formulation	54

1.2	Problem modeling	56
1.3	The model	58
1.4	Classes	60
2	Column-Generation for the LPP-SL	62
2.1	Solution methodology	62
2.2	Pricing problems	64
2.2.1	Bus lines	64
2.2.2	Passenger paths	67
2.3	Column enumeration step	68
3	Computational results on instances from the literature	69
3.1	Comparison with literature results	69
3.1.1	Dutch instance	70
3.1.2	Sioux-Falls instance	71
3.2	Generating bus networks with CGEP	72
3.2.1	Dutch instance	72
3.2.2	Sioux-Fall instance	73
4	Conclusion	74
4	The Frequency Setting Problem (FSP)	77
1	The Frequency Setting Problem	78
1.1	Problem formulation	78
1.2	The model	84
2	Solving process	87
2.1	Generating a set of passenger paths	87
2.2	Passenger Path Selection Process	89
3	Experimentation using instances from the literature	91
3.1	Experimentation settings	91
3.2	Application of PPSP	92
3.3	Frequencies Details for Dutch network	94
3.4	Frequencies details for Sioux-Falls network	94
4	Conclusion	95
5	Case Study	97
1	Construction of the case study	98
1.1	Creating the network model	99

1.2	Generating OD Matrices	102
1.3	Categorizing the selected major stations	104
2	Filtering unrealistic lines	106
2.1	Topology-based filtering rule	106
2.1.1	Major connection lines	107
2.1.2	Meshing connection lines	107
2.1.3	Major diffusion lines	108
2.1.4	Secondary diffusion lines	109
2.2	Length and sinuosity based filtering rules	109
3	Restructuring the Vitalis Network	112
3.1	Bus lines	112
3.2	Service level evaluation	113
4	Designing bus networks with CGEP	115
4.1	Experimentation framework	115
4.2	Estimation of the maximum bus load parameter	116
4.3	Defining the adjustment coefficient	117
4.4	Five propositions of network reoptimization	118
4.4.1	Impact of variations of the maximum-spt-deviation (α) service level parameter for a fixed minimum-rdc-percentage (β) service level	119
4.4.2	Impact of variations of the minimum-rdc-percentage (β) service level parameter for a fixed maximum-spt-deviation (α) service level	120
4.5	Synthesis	122
5	Setting frequencies using PPSP	122
5.1	Experimentation framework	123
5.2	Operator data used in our experiments	123
5.3	Setting the bus filling rate parameter	124
5.4	Setting reference travel time deviation for passengers	125
5.5	Setting up the experimental framework	126
5.5.1	Two hypothesis on passengers behavior	126
5.5.2	Limitations of frequency difference in both directions of a line	127
5.5.3	Bus traffic restriction	127

5.6	Setting the frequencies	127
5.7	Synthesis	131
Conclusion		133
Appendices		137
A	Line Planning Problem	137
A.1	Overestimation of the number of direct travelers with DC constraints	137
A.2	LPP-SL linearization	139
A.3	LPP-SL master and dual problems	140
A.4	Bus Line Labeling Algorithm	141
A.5	Passenger Path Labeling Algorithm	145
B	Frequency Setting Problem	147
B.1	Passenger path requiring to be added to make the solution feasible .	147
C	Literature instances	148
C.1	Dutch instance	148
C.2	Sioux-Falls instance	151
D	Evolution of the bus network of Poitiers	152
D.1	Pre-restructuring network	152
D.2	Post-restructuring network	153
E	Proposed network	154
E.1	Post-restructuring network retained lines	156
E.2	Created Major connection lines	156
E.3	Meshing connexion lines created	159
E.4	Major diffusion lines created	161
F	Line frequencies	163
F.1	Operated line frequencies for the pre-restructuring RTP network . .	164
F.2	Operated line frequencies for the post-restructuring RTP network .	165
F.3	PPSP-proposed line frequencies for the post-restructuring RTP net- work	166
F.4	PPSP-proposed line frequencies for the CGEP-proposed network <i>B</i>	167
Bibliography		169

INTRODUCTION

As reported by the United Nations WUP [2018], approximately 55.3% of the world's population lives in urban agglomerations in 2018 and this percentage will reach 60% in 2030. In addition, one in three people will live in a city with at least half a million inhabitants. Ritchie and Roser [2018] state that the world's population will be 9.8 billion by 2050 and that 68% of people will live in cities. It is therefore a major challenge to design efficient transit networks in urban areas in order to be able to adapt to this large increase in urban population.

The development of an efficient urban transportation network, or the optimization of an existing network, can be achieved at several levels. First, stop locations and street sections can be defined to accommodate new neighborhoods. Second, the lines of the network can be improved or even redesigned in response to major changes in passenger flows. Then, the frequencies of each line can also be revised for different periods of the day to meet the passenger demand. Finally, bus schedules can be adjusted to better serve passengers with transfers at interchange points. This thesis intend to contribute to this strategical topic with optimization tools that can support decision at several level in the design and optimization of an urban transportation network.

This introductory chapter is divided into four sections. The first section focuses on the evolution of urban public transport from its introduction in the mid-19th century to the present days with a description of current issues for network managers. A description of the characteristics of the networks is also given. The second section focuses on the positioning of the thesis in the global process of bus planning. The objectives of the thesis are clearly defined in the third section. Finally, the fourth section lists the contributions made in this thesis.

1 History of urban transportation networks

The information presented in this section is based on the book of Aguilera-Belanger et al. [2017]. Urban public transport is a system of transport for passengers by group travel systems available for use by the general public. They are carried out within the

territorial jurisdiction of a mobility organizing authority, called AOM in France.

Urban public transport has evolved considerably over the course of history. The first urban public transport services appeared in the second half of the 19th century. At that time, they were built and usually operated by private operators under long-term concessions. To offset the risks involved, these operators relied on the steady growth of the urban population. These public transport services were thus profitable for their operators until the end of the first half of the 20th century.

The 1950s was the golden age of the affordable car. This led to a decline in public transport ridership and thus a deterioration in the revenues of the operators. Operators reacted by raising fares, which resulted in more people turning away from this mode of transport. In order to reduce operating costs, services were redefined to reduce wages as much as possible. As these measures were not sufficient, the vast majority of operators decided to reduce maintenance costs and replace infrastructure, leading to a gradual replacement of streetcars by buses.

Between the 1960s and the 1980s, new residential areas were built, resulting in unprecedented urban development.



Figure 1 – A bus in the streets of Nantes in 1966 (Archives de Nantes - Fonds P.F. Gerard)

To compensate for this, managers extended existing bus lines to serve these new residential areas, without much redesign of their network. Unfortunately, as no investment was made in the road network at that time, buses and cars share the road with each other (as shown on Figure 1). As a result, the bus lines remained stuck in traffic and performance deteriorated rapidly, while operating costs increased for the operators.

In the early 1980s, the concept of public service delegation appeared. Private public transport operators were retained to operate the networks, but the local authority assumed responsibility for the investments, thus providing the operator with all the financial resources needed to operate the network. In return, the delegation contracts are reduced and the municipalities acquire greater power in the operation of the networks. The emergence of this principle of delegation allows new investments to be made. Thus, many transport operators decided to bring back the tramway. The city of Nantes, for example, brought back the tramway in 1985, being one of the first in France (as shown on Figure 2). In addition, investments were made in bus lines, separating buses from car lanes, which made public transport more attractive.



Figure 2 – Inauguration of the Nantes tramway on 7 January 1985 (Archives Semitan)

Between 1999 and 2010, the French Institute of Statistics and Economics, known as INSEE, calculated that the area of France's largest urban areas, based on commuter flows, increased by 39%. These changes in the demography of urban and interurban areas have led to a profound change in the travel flows of inhabitants. Indeed, since the early 2000s,

peri-urban areas, which were previously essentially residential, have seen an increase and diversification of their activities. At the city level, the change in travel patterns is reflected in an increase in the number of trips between interurban areas, which is taking precedence over axial trips from the city center to interurban areas. This increase in inter-city travel therefore leads to an increase in trips with more or less centralized origins and destinations.

The main needs for improvement in terms of public transport provision today, whether for work or other reasons, are therefore in the peri-urban areas. The challenge for public transport is therefore to capture a larger share of the flows in these new urban centers. This capture of the flows can be done through the proposal of an urban transportation network that is adapted to this new demand. The challenge is hence to propose a network that is efficient in terms of costs for managers and qualitative for passengers. The work addressed in this thesis is therefore to propose answers to this challenge.

2 Context of the thesis

Lumiplan is a French company whose activity is divided into three sectors : citizen communication, mobility and dynamisation of ski areas. The group was founded in 1972 with the creation of a light plan to inform tourists of some seaside resorts of the Atlantic coast and then ski resorts. In 1992, Lumiplan launched its Transport activity and participated in the European Union's reflections on information systems for public transport. In 1997, Lumiplan deploys the first network of autonomous solar-powered passenger information terminals in an urban environment (Lyon). In 2002, the company integrated "Full Colour" display solutions into its applications, becoming the first player to deploy outdoor color screens in France, particularly in sports venues. In 2007, Lumiplan wins the tender of the electronic bulletin board of the city of Paris and acquires the Heurès software (France), a solution for public transport companies to manage bus schedules and drivers. In 2009, Lumiplan continued its development in the public transportation market by acquiring Duhamel, a pioneer in onboard communication solutions. At the same time, Lumiplan took its first steps in smartphone applications, and published the first mobile applications for ski resorts in France. Lumiplan India was created in 2014 in Bangalore. This subsidiary is dedicated to transport activities. On the same year, smartphone applications experience a new boom with the launch of citizen applications. In 2015, the Mountain activity intensified its development in the United States with the creation of the subsidiary Lumiplan North America. In 2017, Lumiplan continued to grow

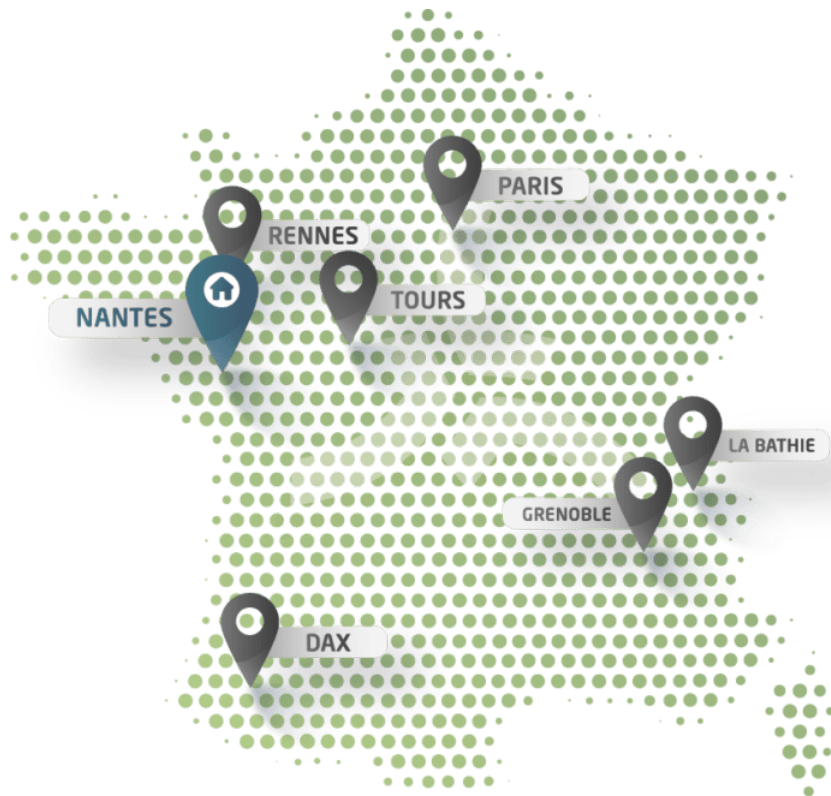


Figure 3 – Rennes (35), Tours (37), Grenoble (38), Dax (40), Nantes (44), La Bâthie (73) et Paris (75)

with the acquisition of Ameli in Italy, a specialist in on-board passenger information for public transport, and Soridis, an expert in passenger information in SNCF stations. In 2019, Lumiplan acquires the Spanish company BQB, a specialist in passenger information solutions for public transport in Spain. Finally, in 2022, Lumiplan will strengthen its expertise in mobility data with the acquisition of Okina.

Currently, Lumiplan makes nearly 50M€ in sales, of which 25% are realized in exports, and hires 270 employees on its 7 operational sites in France (as shown on Figure 3). Lumiplan includes the three following entities:

- *Smart City*, which produce urban signage and develop urban applications,
- *Smart Mobility*, which concerns the mobility part,
- *Smart Resort*, which propose communication panels and applications for ski resorts.

Figure 4 describes the distribution of the revenues between these three entities.

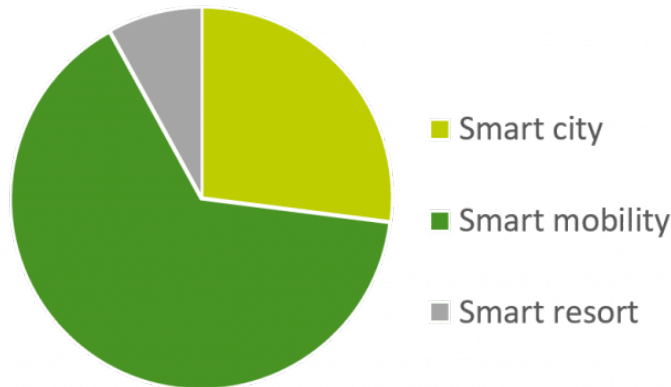


Figure 4 – Repartition of revenues for the year 2018

This thesis takes place in the second entity: Smart Mobility. In addition to orienting and guiding passengers during their journeys, this holding aims to optimize the offer and mastering the mobility data by proposing three softwares: MobiRef, MobiFlux and Heurès.

- MobiRef is a software that build and manage all mobility data of a multimodal network.
- MobiFlux is a software which support the publication of information for travelers in real time.
- Heurès allows vehicle and driver scheduling and schedule management. Figure 5 provides an overview of Heurès. Vehicle scheduling involves designing a rationalized offer by constructing the vehicle services while driver scheduling involves constructing the workdays required for operations. This part was co-developed by Lumiplan in partnership with EURODECISION, a company specialized in optimization. Finally, the schedule management modules support the user in the management of the resources such as buses and Drivers that are available on a daily basis.

Many cities, in France and elsewhere, use Heurès to manage their bus network. For example, this is the case of Angoulême, La Rochelle, Poitiers, Troyes, Saint Nazaire, Algiers, Madrid, Mexico City and Rabat.

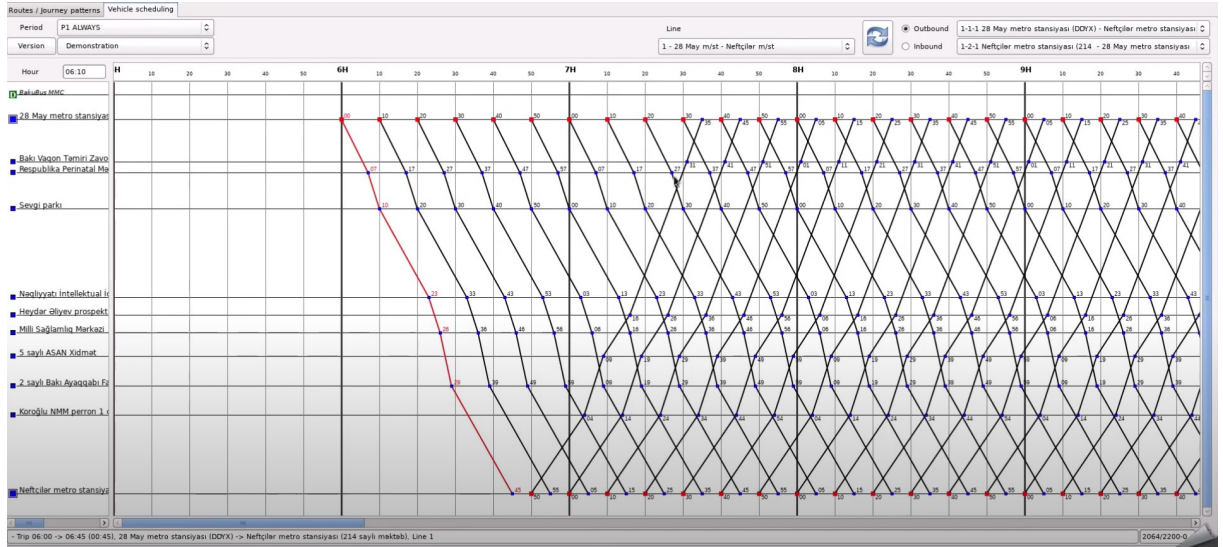


Figure 5 – Example of vehicle scheduling with the Heurès software

3 Objective of the thesis

Lumiplan currently offers Heurès software to public transport companies to support their vehicle and driver scheduling operations. Based on an industrial partnership (CIFRE) between IMT Atlantique, the LS2N laboratory and Lumiplan, this work is part of Lumiplan’s objective to offer, in addition to the Heurès software, a tool to provide scenarios for more efficient public transportation offers. The design of a public transportation network must be addressed by integrating multiple issues:

- Cost (commercial and total operated kilometers, vehicle fleet, ...)
- Quality of service (adherence to passengers’ needs in terms of origin and destination; travel time, directness of the trips, comfort on board, ...)
- Environmental impact (CO₂, NO_x, ...)

This thesis focuses on the design of bus networks through the construction of bus lines and the allocation of frequencies to these lines. Hence, the first part of the thesis deals with the construction of the network, through the determination of a set of lines to operate. This step is achieved by solving the LPP. In this step, the operating costs are set as an objective and some service-levels are incorporated to model passenger comfort. The second part of the thesis focuses on the determination of line frequencies. The associated problem for this step is the FSP. In this step, operating costs are set as an objective and

some shortest path constraints are incorporated to model passenger choices based on the frequencies assigned to the lines. This problem can be solved after the construction of the network or regularly during possible updates of the travelers' requests.

One of the expectations of Lumiplan was the scalability of the proposed models. Indeed, the possibility to use an origin-destination matrix containing transit ridership data as well as the possibility to use an origin-destination matrix for all modes of transportation was something desired. This possibility allows public transportation operators to offer a better quality network to a group of users but also to capture the demand of travelers who do not yet use public transportation networks. This flexibility is also expected in terms of infrastructure, so that our models can work with street sections already used by buses and with street sections that are not yet used. Figure 6 describes the use of this scalability through the different stages of the mobility conception.

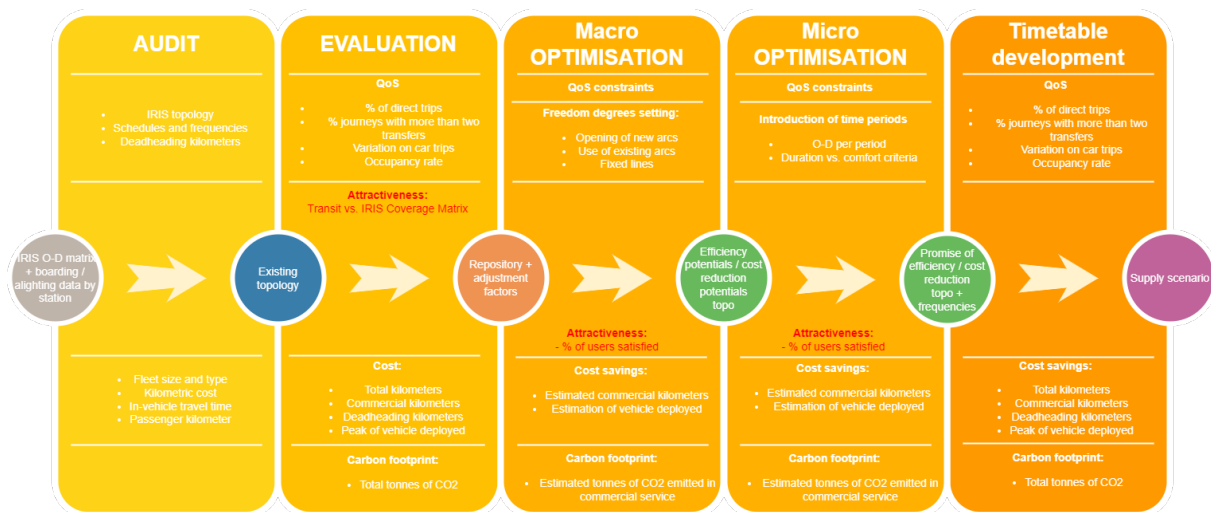


Figure 6 – Stages of mobility conception

The ability to optimize an urban transportation network is intended to be adapted according to the demands of public transportation operators. Indeed, some operators prefer to modify their network from the strategic level while others will prefer to modify their network only at the tactical level. This results in five use cases for urban transportation network operators:

1. Complete redesign of lines and frequencies
2. Partial redesign of lines and complete redesign of frequencies
3. Partial redesign of lines and redesign of frequencies on new lines only

4. Fixed lines and complete redesign of frequencies
5. Fixed lines and partial redesign of frequencies

Figure 7 summarizes the prospects for optimizing public transportation networks using the contributions of this thesis.

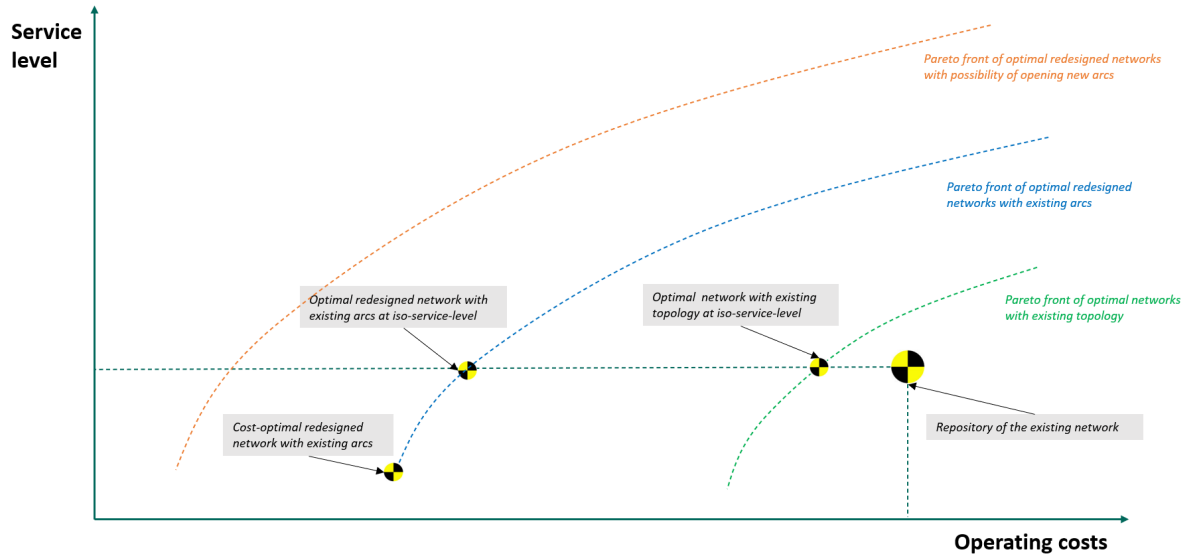


Figure 7 – Applications of the models and algorithms of the thesis to public transportation networks optimization applications.

4 Organization of the thesis

This thesis manuscript consists of four chapters defined as follows:

- Chapter 1 presents the overall bus networks planning process. A particular focus is given to the two steps of constructing the line plan and determining the frequencies of the lines. In the literature, the two problems associated with this step are the line planning and the frequency setting problems. We provide a review of the literature dealing with these two problems. Finally, we present the most common service level criteria, taken either from the scientific literature or from public legislation.
- Chapter 2 discusses the methodology chosen to decompose the general problem into two steps: construction of the lines plan and determination of the frequencies of the lines.

- Chapter 3 is devoted to the LPP. We first introduce a model, issued of the literature and reinforced by specifying operational constraints and service level parameters for the passengers, is described. A resolution process based on a column generation method is also presented. Some associated results from literature instances are then described.
- Chapter 4 is devoted to the FSP. A single-level linear model that minimizes the operator's costs while respecting shortest paths constraints for travelers is presented. A resolution process based on a dynamic selection of passenger paths is also described. Finally, some results from literature instances are described.
- Chapter 5 is devoted to the case study based on the network of the agglomeration of Poitiers. This chapter highlights the interest of our work in the context of a re-optimization of an urban transportation network.

Chapter 3 is the subject of a paper under finalization to be submitted soon in an international journal. The results concerning this chapter have been presented in the conferences [Gatt et al., 2021] and [Gatt et al., 2022b]. Concerning Chapter 4, some results were presented in the conference [Gatt et al., 2022a].

LITERATURE REVIEW

Contents

1	The global planning process	28
2	The Line Planning Problem	29
2.1	Definition	29
2.2	Solving approaches	32
2.3	Service levels	37
3	The Frequency Setting Problem	40
3.1	Definition	40
3.2	Methods	41
4	Research gaps	44
4.1	Line Planning Problem	45
4.2	Frequency Setting Problem	45

This chapter is devoted to the study of the planning of urban transportation networks at strategic and tactical levels. At these levels, the planning of urban transportation networks translates into the two steps of bus network design and frequency setting.

The first section focuses on the presentation of the global planning process. The second section describes the line planning problem, which consists in defining the set of lines to be operated. In this section, two solving approaches are described and some service levels are investigated. The third section describes the frequency setting problem, which consists in specifying the frequencies of the lines for a set of considered time periods. Finally the fourth section investigates the research gaps of these two problems.

1 The global planning process

According to Ceder and Wilson [1986], there are generally five steps in the planning process of transportation system. These steps, described in Figure 1.1, range from long-term strategic decisions to short-term operational decisions. At the strategic level, the first step is the construction of the network. This stage itself has several aspects: creation of stops, creation of new sections, design of bus lines to serve the bus stops and bus sections. In our work we focus on the determination of the lines assuming the stops and sections already exist. Halfway between the strategic and the tactical level is the step of frequency determination. This step can be carried out after the construction of the network or at regular intervals in order to adapt to variations in passenger demand. The determination of schedules and the creation of bus and driver schedules are all three operational-level steps and are performed on a seasonal basis to adapt schedules to demand fluctuations.

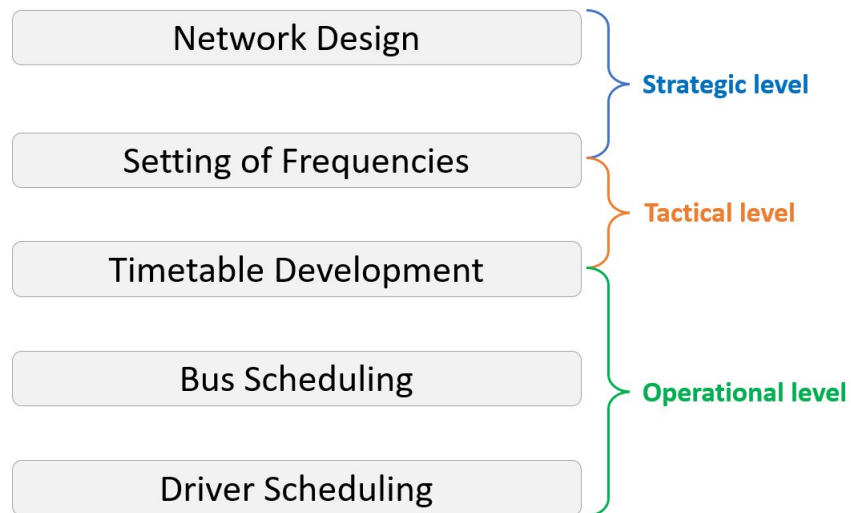


Figure 1.1 – Global process of bus network planning according to Ceder and Wilson [1986]

In our work, we focus on the first two steps of the bus network planning process: the design of the bus network, obtained by solving the Line Planning Problem (LPP) and the setting of the frequencies of the lines, obtained by solving the Frequency Setting Problem (FSP).

2 The Line Planning Problem

This section presents a review of the literature on the line planning problem. A definition of the problem is first given and an analysis of the solution approaches is then made. Finally, we provide an analysis of the service level parameters considered by the literature in optimization models.

2.1 Definition

The Line Planning Problem (LPP) [Borndörfer et al., 2007, Borndörfer and Karbstein, 2012, Schöbel, 2012, Karbstein, 2013, Goerigk and Schmidt, 2017], also called Transit Network Design Problem (TNDP) [Cipriani et al., 2006, Guihaire and Hao, 2008, Mauttone and Urquhart, 2009, Cipriani et al., 2012, Liang et al., 2019] or Transit Network Design and Frequency Setting Problem (TNDFSP) [Nikolić and Teodorović, 2014, Arbex and da Cunha, 2015, Jha et al., 2019, Durán-Micco and Vansteenwegen, 2021], consists of determining a set of bus lines to operate. Some papers consider that the LPP also consists of assigning a frequency to each bus line. This may be necessary to better estimate the network cost or demand satisfaction. The decisions resulting from solving the LPP fall into the class of strategic or long term decisions in the planning process. The inputs of this problem are the infrastructure network and the demand matrix [Durán-Micco and Vansteenwegen, 2021].

The infrastructure network used to solve the LPP is represented by a public transportation network composed of bus stops and street sections. In our case we note $G = (V, A)$ the graph that models this network. The set of nodes V represents the bus stops and the set of arcs A the street sections connecting the bus stops. Each street section is oriented and is characterized by a distance, a in-vehicle travel time and a maximum capacity. In order not to generate a too large graph, the nodes considered for the LPP are defined from the major stations.

As defined by Karbstein [2013], given a set of termini $T \subseteq V$, a line is a simple path in the urban transportation network $G = (V, E)$ that starts and ends at a terminus. At the LPP level, it is assumed that lines are operated in both directions and with homogeneous vehicles.

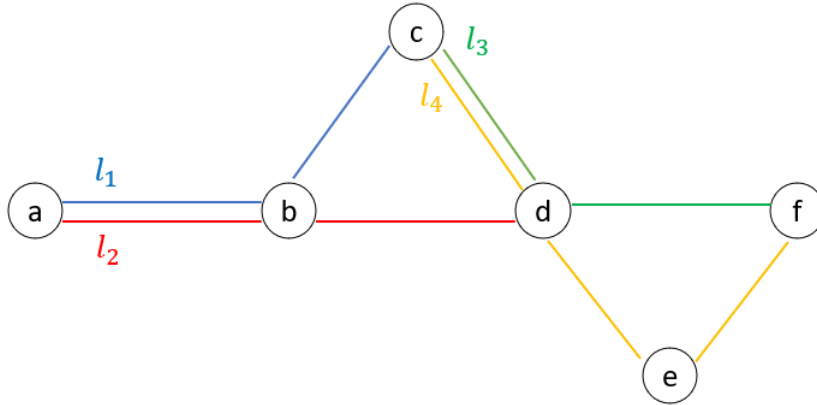


Figure 1.2 – Miniature transportation network with six bus stops and four bus lines: $l_1 = (a, b, c)$, $l_2 = (a, b, d)$, $l_3 = (c, d, f)$, $l_4 = (c, d, e)$

Each line $l \in \mathcal{L}$ is associated with a frequency f_l between a minimum and a maximum value. This frequency describes the number of bus services on the line during a given period of time.

The demand matrix is called *origin-destination matrix* or *OD-matrix*. It contains the so-called *OD pairs* that are defined on the nodes of the network. An OD pair models

$$D = \begin{pmatrix} q_{(a,a)} & q_{(a,b)} & q_{(a,c)} \\ q_{(b,a)} & q_{(b,b)} & q_{(b,c)} \\ q_{(c,a)} & q_{(c,b)} & q_{(c,c)} \end{pmatrix}$$

Figure 1.3 – Example of an OD matrix for a network composed of three stations $\{a, b, c\}$

the quantity of passengers that are expected to travel from an origin to a destination on the *planning period*. All origins and destinations being defined on the nodes of the network. Figure 1.3 shows an example of an OD matrix on a 3-nodes network. $q_{(a,b)}$ is the OD pair defining the number of users desiring to go from stop a to stop b . Since a bus network remains the same at the scale of an operating day, the problem is solved over a large time period with a large amount of OD pairs. To integrate this, Ceder and Wilson [1986] propose to define the OD matrix for the major distinct periods of operation such as morning peak, midday, afternoon peak or evening. Baaj and Mahmassani [1995], Durán-Micco et al. [2022] constitute their OD matrix with the demands of the morning peak period, which is the time period in the day where they observe the highest demand. For

their part, Zhao and Ubaka [2004] use an OD matrix which represent the daily number of passenger trips between each pair of zone centroids for the area of Miami-Dade. Finally, in a similar way, Mumford [2013] propose to use an OD matrix composed of average values in a 12 hours time horizon, to represent the passenger need across all the operation period. Thus, we observe two types of OD matrices used for LPP, the first composed of daily peak demands and the second composed of the sum of daily demands.

A passenger path is associated to a given OD pair and is defined as a sequence of edges relying the origin node of the OD pair to its destination node. On Figure 1.2, the two possible paths for a (a, c) OD pair are $p_1 = [(a, b), (b, c)]$ or $p_2 = [(a, b), (b, d), (d, c)]$.

Regarding objective functions, we distinguish two perspectives on the problem. The first one is from the passenger’s point of view while the second is from the operator’s point of view.

For the passenger’s point of view, most approaches [Mandl, 1980, Ceder and Wilson, 1986] minimize a sum composed of the in-vehicle travel time and a fixed penalty for transfers. Some other paper propose to maximize the total demand served [Bertsimas et al., 2021]. Papers dealing with this point of view, mostly limit operator costs by using budget constraints. Some other studies propose to minimize the unsatisfied demand [Cipriani et al., 2012], to maximize the number of direct passengers [Bussieck et al., 1997, Suman and Bolia, 2019] or even to maximize the transit trip density [Yu et al., 2012].

Regarding articles which consider the operator’s point of view, the operating costs are generally considered as the objective to be minimized. These operating costs are directly related to the number of kilometers of operated lines. According to Schöbel [2012], the most common approximation of the operating costs of a line concept (\mathcal{L}, f) where \mathcal{L} is the set of lines to operate and f their associated frequencies, is defined as:

$$c(\mathcal{L}, f) = \sum_{l \in \mathcal{L}} cost_l \times f_l \quad (1.1)$$

where $cost_l$ describes the costs for operating one vehicle along line l during the planning period and f_l the frequency of the line l .

Concerning the service level in papers dealing with this approach, we can observe some limitation on in-vehicle travel times of passengers [Suman and Bolia, 2019] as well as limitations on the maximum number of allowed transfers [Yu et al., 2012].

Many studies propose to aggregate these two views as a weighted sum [Borndörfer et al., 2007, Marín and Jaramillo, 2009, Borndörfer and Karbstein, 2012, Camporeale

et al., 2019], while others propose to model the LPP as a bi-objective model to find solutions with different trade-off between cost and passenger service [Mumford, 2013]. Finally, some studies propose to integrate the CO2 emission as an objective to be minimized [Duran-Micco et al., 2020].

2.2 Solving approaches

According to Karbstein [2013], there are two approaches to model the choice of paths made by passengers. The first approach is called "Fixed Passenger Routing" and considers the passenger's choice of paths as independent of the available set of lines. This approach, used by Bouma and Oltrogge [1994], Bussieck et al. [1997], assumes that the passenger's path choice is known before solving the problem and cannot be questioned by the new generated lines. The second approach is called "Integrated Passenger Routing". It considers the passenger choice as line-dependent. Thus, as new bus lines are generated, the passenger's choice of path can be modified. Schöbel and Scholl [2006] show that, in the case of the "Integrated Passenger Routing" approach, the line planning problem with budget constraints is NP-hard. In our work and in this literature review we focus on studies using this "Integrated Passenger Routing" approach.

Regarding the creation of new lines, two approaches can be distinguished. The first approach consists in separating the LPP resolution in two distinct steps: the first step generates a pool of bus lines, called *line pool*; and the second step selects lines in this a priori generated line pool. We call this type of approach *a priori line pool generation approach*. The second approach consists in generating the bus lines while solving the LPP. We call this type of approach *integrated line generation approaches*. Concerning this second type of approach, Heinrich et al. [2022] consider the complexity of the LPP when all simple paths can be used as lines. They show that the problem is NP hard and that no polynomial time resolution is possible except for trees graph when $f^{min} = f^{max}$. The LPP literature is detailed below for each of these approaches.

2.2.1 A priori line pool generation approaches

Studies using a priori line pool generation approaches require using a qualitative line pool. The generation of a line pool can be done in different ways. Ceder and Wilson [1986] hence propose a method to generate a line pool by enumerating all lines with lengths that do not exceed a factor of the shortest path length between their terminus. More recently,

Gattermann et al. [2017] propose a method to generate a line pool by using the results of a minimum spanning tree problem on the transit network. Among all these studies using a priori line pool generation, some of them are based on MIP approaches while other studies are based on heuristic methods.

Heuristic approaches

Concerning studies using heuristic, Mandl [1980] propose a method starting from a set of feasible lines that is heuristically improved with a local search method based on three improvement possibilities: (1) exchanging parts of the lines, (2) adding a station to the lines, (3) removing a station from the lines. This local search method is executed with the aim to minimize the total in-vehicle travel and transfer times of passengers. Baaj and Mahmassani [1990, 1995] propose to generate bus lines connecting high-demand node pairs along either their shortest path in terms of the total in-vehicle travel time or along alternate short path whose round trip time is within a threshold of the shortest path's travel time. They then propose to use an iterative process called TRUST in which each bus line of the line pool is assigned a frequency. Frequencies and passenger flows are then updated sequentially until the value of frequencies converges (although the process does not guarantee convergence). Cipriani et al. [2006] first generate a line pool and then use a genetic algorithm to solve the LPP. Their objective is to minimize the sum of operator's costs, passengers' costs and external costs. The authors test their algorithm on an instance based on a neighborhood in Rome, Italy, consisting of 49 nodes and with OD pairs referring to public transportation and private cars. Cipriani et al. [2012] extend this method to solve a large instance based on a real case and composed of 1300 nodes. In this case study, the number of OD pairs is not specified. However, compared to Cipriani et al. [2006], they simplify the OD pairs to refer only to public transportation. Mumford [2013] propose an evolutionary bi-objective framework, with the objectives of minimizing operator cost and passengers travel time, incorporating in-vehicle travel time and penalties for transfers. A heuristic-based method is proposed to generate an a priori line pool. Their method is based on four steps : generation of an initial line pool, crossover step, repair step and mutation step by using mutation operators to add and delete groups of nodes to the lines. They test their method on the 15 nodes Mandl instance introduced by Mandl [1980]. They obtain better results than other studies for passenger costs when testing it with a number of lines operated set to 4, 6, 7 and 8. Furthermore, they propose four new instances with 30, 70, 110 and 127 nodes and specific number of lines. Camporeale et al. [2019] propose to generate a line pool by using a so-called *candidate route set generation*

procedure and to solve the LPP using a genetic algorithm with this line pool as input. Their *candidate route set generation procedure* is described in Camporeale et al. [2017] and is based on two steps: (1) generation of all possible lines between the node pairs (2) filtering of the lines according to their feasibility. Their genetic algorithm allows to determine a set of lines to operate, each of them associated to a frequency, while respecting some operational constraints such as the maximum fleet size.

MIP approaches

Concerning studies using MIP approaches, many of them are based on column-generation methods, which is a well known solution approach for transportation problems [Desaulniers et al., 2006, Feillet, 2010]. Borndörfer and Karbstein [2012] propose a model minimizing a weighted sum of operating costs and passenger travel times, including in-vehicle travel times and transfer penalties. They separate the passenger paths into two sets: one set of direct paths and the other of indirect passenger paths (e.g., paths with at least one transfer). This specification of the directness of a path, but not of its exact number of transfers, allows to solve the problem for real cases with an acceptable solving time. Moreover, they introduce a new set of passenger path called *relaxed direct connection passenger path* that we abbreviate in *rdc-passenger-path*. A passenger path is said to be a rdc-passenger-path if there exists, on each of its arcs, at least one line $l \in \mathcal{L}$ that circulates on it and goes from the origin to the destination of the path. The formulation of their model with this set allows to eliminate the assignment of direct paths to particular lines. The proposed method uses a line pool which is generated a priori, but it uses column generation to generate rdc and indirect-passenger-paths while solving the model. The line pool is generated with the method of Ceder and Wilson [1986].

Suman and Bolia [2019] propose a method based on two mathematical formulations solved with a commercial solver to optimize the redesign of existing bus lines. Their first formulation focuses on modifying the existing lines. They reuse the method defined by Ceder and Wilson [1986] and enumerate, for each existing bus line, all lines relying the same terminus and with lengths that do not exceed a factor of the shortest path length between the two terminus. The second formulation determines the set of lines that maximizes the total length of passenger trips satisfied without transfer (called direct passenger-kilometers) In this formulation, they integrate operational constraints on the in-vehicle travel time, line length constraints or even constraints on the ability of each bus to operate on the different lines. They test their method on the Delhi area (India) using CPLEX and MATLAB for computations and analyses. The results show a potential gain

in direct passenger-kilometers of more than 35%.

What can be observed from the literature is that a limitation of the a priori line pool generation approach, is that the line pool cannot be too large for the optimization problem to remain tractable. Hence, the resulting solution is highly dependent on the generated line pool, which is necessarily incomplete for problems of realistic sizes.

2.2.2 Integrated line generation approaches

Studies dealing with integrated line generation approaches can be based on two different line generation methods. Some of them propose to generate lines during the resolution of the LPP while others only propose to integrate some of the lines included in an a priori generated pool in order to improve the solution.

Heuristic approaches

Yu et al. [2012] extend the direct traveler density model proposed by [Yang et al., 2007] in order to model a transit network design problem with an objective to maximize traveler density which is linked to the number of passengers carried by the lines of the network. They use an ant colony optimization (ACO) to solve the model. This ACO is based on four steps: (1) generation of a bus line, (2) local search using a 2-opt exchange, (3) passenger assignment, (4) pheromone information update. They test their ACO method on a network based on the city of Dalian (China). This network is composed of 1500 nodes. The OD matrix is based on the peak-hour period and is obtained from Yang et al. [2007]. The obtained results show that the number of transfers can be improve in the transit line network thanks to optimization.

Based on Duran-Micco et al. [2020], Durán-Micco et al. [2022] propose a bi-objective memetic algorithm for the Transit Network Design and Frequency Setting Problem (TNDSP). The two objectives of the problem are: the average passenger travel time and the fleet size. They define the average passenger travel time as the sum of in-vehicle travel time, waiting time, transfer penalization and unsatisfied demand penalization. This algorithm is fed by an initial population generated by a random heuristic selecting lines from a line pool, including the k shortest paths between OD pairs with the highest demand. Their memetic algorithm then combines an iterative process of population renewal using crossover, mutation, and local search operator. An iterative heuristic solves the passenger assignment and the frequency setting problems. The passenger assignment problem is based on Baaj and Mahmassani [1990]. It considers that passengers have a set of attractive lines to reach their destination, which is composed by all the paths that are at most

10% larger than the shortest option, considering only the in-vehicle travel time and the penalization per transfer. They restrict their problem to covering requests with at most one transfer and test their algorithm on an instance, based on the city of Utrecht (The Netherlands), composed of 271 nodes, 470 arcs and 16823 non-zero OD pairs. With their method, they generate solutions with lower average travel times and fleet sizes compared to the current network.

MIP approaches

Among LPP studies based on column generation, Borndörfer et al. [2007] propose a model for the LPP. Their model is designed to dynamically generate passenger paths and bus lines, the flow of both of them being modeled with a path-based formulation. The objective function is a weighted sum of passenger in-vehicle travel times and operating costs. Their method is tested on an instance based on the network of Potsdam (Germany) and composed of 410 stops and 4685 non-zeros OD pairs. The major limitation of their method is that passenger transfers are not taken into account. This leads to unlimited number of transfers and travel times.

Jin et al. [2016] investigate the design of a network of temporary bus services in response to local disruptions in urban rail transit systems. This problem involves a limited total number of lines, limited lines per station, limited fleet, and the need to consider transfer times. They propose a three steps approach: (1) a column generation based procedure is used to dynamically generate bus lines based on the OD pairs, (2) a path-based multi-commodity network flow model is used to identify the most effective combination of the candidate lines to operate, and (3) another optimization based procedure determines the optimal allocation of vehicles among the selected lines and the corresponding headways. For the first step, the objective function of the master problem minimizes the total increase in journey time of all passengers groups over the entire disrupted period. The waiting times, when transferring to buses or rail, are not considered in their restricted master problem. For the second step, the objective is to minimize the path durations (consisting of in-vehicle travel time, transfer time and waiting time) of all passengers. They test their method on a network composed of 109 nodes and 236 arcs. They study two disruption cases and quickly obtain interesting results leading to reductions in the expected path duration delay of system users while respecting operational constraints.

Redesigning a bus network in the presence of a metro system is a variant of the LPP. Liang et al. [2019] consider this problem and propose to solve it by a two-step algorithm. The problem consists in designing a bus transit network that can be coordinated with

the metro network to achieve a balanced modal split between metro and bus. The bus lines and the passenger paths are generated using a column generation process before solving a stochastic linear programming model to optimize the bus line frequencies under demand and bus travel time uncertainty. In this stochastic model, the expected costs of the system, including operator and passenger costs, are minimized. The authors apply their method to the redesign of the Beijing Second Ring public transit network which is composed of 14 interchange metro stations and 38 major bus stops.

Recently, Bertsimas et al. [2021], Yan [2020] propose a column generation approach to generate bus lines while modeling the paths taken by passengers as flows on the network. Their model maximizes the satisfaction of the passenger demand with constraints to limit the network cost. Similar to Borndörfer and Karbstein [2012], direct and indirect passenger trips are distinguished and line frequencies are chosen from a finite set to avoid associating lines with too low frequencies. This method is experimented on a Boston area based network consisting of 233 nodes and 2042 arcs and with an associated OD matrix consisting of 534 OD pairs with only demands of at least 50 passengers during the entire day. 9 lines of the existing bus network are given as input and the column generation based algorithm is run for twenty iterations. 40 bus lines are generated at the end of the iterations, which is very few considering the size of the considered instance. However, even with this small set of lines generated, the resulting networks improves the percentage of served demand by up to 35% compared to the original network.

Compared to the LPP papers following the a priori line pool generation approaches, we notice that there are more papers in the integrated line generation category that are based on MIP approaches.

2.3 Service levels

Service level evaluates the performance of a bus network from the passengers point of view. It can be addressed in various ways. Suman et al. [2018] analyze the perceptions of potential passengers of existing bus services in Delhi, India, and conclude that people do not use buses because the bus lines are not direct enough, the buses are overcrowded, using them takes too much time compared to other transportation modes and punctuality is not good.

The European decree EN-, published in September 2002 and still operating, specifies the requirements for defining, targeting and measuring the quality of service of a transportation network. It defines two performance measures at the network design scale. The

first one is called "possibility to go wherever, whenever". It implies that the stops in the network can be connected to each other within a predefined time or distance limit. The second is called "convenience". It implies that it is possible to travel with a minimum number of transfers on the network. A quantification of this performance is proposed by setting a minimum number of customers having direct journeys. We now propose to study the use of these two performance measures in the LPP literature.

2.3.1 Travel time related service levels

A first approach to evaluate the service level provided by a network is to analyze the travel time of each passenger on this network. Some studies propose to minimize passenger travel times composed of in-vehicle travel times and to include penalties for transfers [Schöbel and Scholl, 2006, Mumford, 2013, Jin et al., 2016]. Schöbel and Scholl [2006] hence construct a so-called "change-and-go" graph containing copies of arcs for all lines and additional arcs for each possible transfer of passengers from one line to another. The resulting graph is very dense as the number of lines increases. Models based on this graph can be solved only for small networks.

Following a different approach, some papers integrate passenger travel times as a constraint. Thus, some studies enforce all passenger paths to have a duration constrained by a maximum deviation to the shortest in-vehicle travel time of passengers. Ceder and Wilson [1986] and Baaj and Mahmassani [1995] both enforce passengers not to use a bus line if the resulting trip exceeds a maximum factor of its shortest feasible in-vehicle travel time. Ceder and Wilson [1986] set this parameter to 50% while Baaj and Mahmassani [1995] propose to set it to 40%.

Yan [2020] and Bertsimas et al. [2021] propose two models which objective are to maximize passenger satisfaction with a maximum network cost limit. In the first model, passengers can only be assigned to a zero-transfer trip, while in the second model, passengers can be assigned to a zero-transfer trip or a single-transfer trip with no preference. In both model, a maximum deviation to shortest path is used. The authors propose to set this parameter to a value of 50% in their experiments.

Durán-Micco et al. [2022] reuse the maximum deviation to shortest path by using a passenger assignment problem based on Baaj and Mahmassani [1990]. In their passenger assignment problem, passengers have a set of attractive paths to reach their destination at most 10% larger than the shortest option. In this study, the travel time of passengers includes the in-vehicle travel time and potential transfer penalties.

Finally, some studies prefer to set a travel time limit to concurrence the private transport mode. Marín and Jaramillo [2009] propose a model for the urban rapid transit network design problem that they define in Marín [2007] as finding the optimal lines and stations in an urban traffic context. Their objective is to maximize the public transportation demand satisfied by the new infrastructure while respecting constraints on budget and lines. The problem has two decision levels: first, the system operator selects the stations to be served from a set of potential stations and with the objective of maximizing demand coverage. Then, the travelers are assigned to the network by solving a multi-modal assignment problem. The authors introduce a constraint enforcing OD pairs to be assigned to the public transit mode only if the travel time by using public transport is less than the private mode travel time. The problem is solved by Benders decomposition. The algorithm is tested on the bus network of the city of Seville (Spain) which includes 24 nodes, 264 edges and 552 OD pairs. Gutiérrez-Jarpa et al. [2017] present a discrete, multi-objective mathematical program for strategic planning of a rapid transit system. The objectives are to maximize the total flow captured by the rapid transit system compared to the car mode, the total time savings of the captured flow with respect to car travel times, and to minimize the total construction cost of the system. An origin-destination flow is considered captured only if the estimated time required for the associated passenger to use the network is at most equal to its travel time by car.

2.3.2 Transfers related service levels

The concept of directness of a network refers to the percentage of travelers with a direct path (ie. without transfer). For the same travel time, a direct trip is always preferred to a trip that requires one or more transfers. Therefore, the possibility for a traveler to take a direct path is a major asset in terms of attractiveness for a network.

Some studies decide to set a constraint on the maximum number of transfers on a passenger path. Yu et al. [2012], Duran-Micco et al. [2020], Bertsimas et al. [2021], Durán-Micco et al. [2022] hence limit the number of transfers to one. Nevertheless, this limitation to zero-transfer and one-transfer passenger paths may be restrictive in some cases where passengers would accept two transfers. Baaj and Mahmassani [1990, 1995], Mumford [2013] limit the maximum number of transfers to two and assume that a trip with more than two transfers is unacceptable.

Many studies propose to integrate a penalty per transfer in their model. Among the papers implementing this, we find Mandl [1980], Ceder and Wilson [1986], Baaj and

Mahmassani [1990, 1995], Yu et al. [2012], Mumford [2013], Jin et al. [2016]. Among them, Baaj and Mahmassani [1990, 1995], Mumford [2013] propose to set this transfer penalty to five minutes. Nevertheless, when it is combined with travel time constraints or with a passenger travel time minimization objective, the use of a transfer penalty does not allow to evaluate the number of voyagers having a direct path.

As a result, other studies propose to take the directness of the network as an objective to maximize. Zhao and Ubaka [2004] hence propose a method to solve the line planning problem with the objective to minimize transfers and optimize line directness while maximizing service coverage. Yu et al. [2012] use an ant-colony optimization method to solve the line planning model with an objective to maximize directness and transfer demand density of lines under some resource constraints. In a similar way, Suman and Bolia [2019] maximize the number of kilometers traveled by passengers with a direct path.

Finally, other papers propose to set a minimum percentage of OD pairs with a direct path. Among them, we find Baaj and Mahmassani [1990, 1995] who propose to set a minimum of 50% of passengers to have a direct path.

3 The Frequency Setting Problem

This section review the literature related to the frequency setting problem. A definition of the problem is first given and an analysis of the different models and solving approaches is then made.

3.1 Definition

The Frequency Setting Problem (FSP) determines the frequency at which a given set of lines should be operated in order to ensure a high level of service for a given set of time periods. The frequency of a line can be defined as the number of trips operated during a given time period. Hence, for each line, a frequency is given per direction and per time period. The FSP can be solved during the redesign of the network, after solving the LPP, or on a tactical level, to adapt the line frequencies to evolutions of the passenger demand.

The network infrastructure used to solve the FSP is represented by a public transportation network with given sets of bus stops, street sections and lines. Each street section is oriented and is characterized by a distance, an in-vehicle travel time and a maximum capacity. In our case we denote by $G = (V, A, \mathcal{L})$, the underlying bus network. We

assume that the demand is expressed for pairs of bus stops. The demand is represented using an origin-destination matrix noted D^t defined for each time period $t \in T$ of duration $\Delta(t)$. Compared to the LPP, the bus fleet considered for the FSP is heterogeneous. This heterogeneity may be related to operating cost and bus capacity.

Compared to the LPP, passenger travel time in the FSP incorporates not only vehicle travel time but also waiting time at departures and transfers. Hence, we define the *path duration* of a passenger journey as the sum of its *in-vehicle travel time* and its *expected waiting time*. In addition, the FSP better consider passengers preferences regarding their choice of path. For this purpose, we model a passenger path as a succession of arcs, each associated with a bus line, that connect the origin to the destination of the corresponding OD pair. The user of a public transportation network chooses a path based on his own interest, which is generally expressed by minimizing his travel time. Thus, to ensure the efficiency of the transportation system, the models must take into account the behavior of the users. Thus, a better estimation of how to deploy the bus capacities on the network is provided when the passenger path choices are modeled accurately. The determination of passenger flows on the network is called *transit assignment*.

3.2 Methods

According to Ibarra-Rojas et al. [2015], there are two major types of approaches to solve the frequency setting problem : (1) without transit-assignment decisions and (2) with integrated transit-assignment decisions.

For the first approach, it is assumed that the choice of path made by the passengers on the network, does not depend on the line frequencies and is known in advance. A simple version of the problem consists in determining the frequency of each line given a maximum fleet size such that the sum of all passenger in-vehicle travel times is minimized [Newell, 1971]. Han and Wilson [1982] propose an algorithm to solve the problem of allocating a fleet of buses in a public transportation system, which is a variant of the FSP. Their algorithm first recognizes passenger path choice and then solves a model which minimizes the occupancy level at the most loaded point on any line in the system, subject to bus fleet and capacity constraints.

A major issue of the problem is to model passenger behaviors. Hence, the second approach, which considers transit assignment decisions, integrates this behavior which is reflected in the path choice. Since the duration of a passenger path depends on the expected waiting time at the first stop and at each transfer on this path, it depends on

the frequency assigned to each line on the path.

Accordingly, the FSPs that integrate transit-assignment decisions are mostly formulated as bi-level optimization problems. A bi-level problem is composed of a leader problem called upper-level problem and of a lower-level called follower problem. The decision of the follower problem is parameterized by the leader's choices and these choices are constrained by the follower's decision. Hence, a bi-level model can be expressed as:

$$\min_{x \in X, y} F(x, y) \quad (1.2)$$

$$s.t. \quad G(x, y) \geq 0 \quad (1.3)$$

$$y \in S(x) \quad (1.4)$$

Where (1.2), (1.3) and (1.4) compose the upper-level problem and where S is the set of optimal solutions to the problem.

$$S(x) = \operatorname{argmin}_{y \in Y} \{f(x, y) : g(x, y) \geq 0\} \quad (1.5)$$

The upper level consists of assigning a frequency to each bus line setting the frequencies of the bus lines while the lower level consists of assigning a path to each passenger according to its path duration. The line preference is hence dependent on the waiting times of the bus lines and therefore on the frequencies assigned to these bus lines.

Upper-level problems

Two types of objective can be chosen for the upper level model. The first type is called passenger-oriented and consists of minimizing the sum of all passengers path duration, generally associated with fleet or budget constraints. The second type of objective is called cost-oriented and consists in minimizing the operator's operating costs.

Constantin and Florian [1995], Noriega and Florian [2003], Yu et al. [2010], Martínez et al. [2014], Arizti et al. [2018] choose the passenger-oriented objective for their upper-level and propose to minimize the sum of all passengers path duration.

Some studies propose to mix cost and passenger-oriented components in the objective function of their upper-level problem. Hence, the upper-level function of Ruisanchez et al. [2012] is based on an aggregation of user and operator costs and is subject to technological and demand satisfaction constraints. Giesen et al. [2016] propose a multi-objective bi-level model based on the minimization of the sum of all passengers path duration and on the minimization of the fleet size required for operation. Ma and Chow [2021] chose an

objective based on a weighted sum of operational costs, user waiting costs, user access and egress costs and in-vehicle time costs.

Lower-level problems

Leventhal et al. [1973] propose a column generation algorithm for solving a class of non-linear traffic assignment problems. Few years later, Spiess and Florian [1989] present a transit assignment problem based on the concept of optimal strategies. This model, using a fixed set of bus lines, considers that a traveler chooses his optimal strategy. This strategy consists in choosing the path that allows him to reach his destination at the minimum expected cost. They propose two formulations: the first one where no congestion effect occur and the second one where congestion effects are integrated. With the first formulation, the waiting time at a stop depends only on the frequencies of the lines considered by the traveler. For the formulation, they propose a linear programming problem of a size that increases linearly with the network size. They test their algorithm on the transit networks of Winnipeg, Manitoba, Montreal, Quebec, Stockholm Inner City, Sweden, Oakland, California, and Ottawa, Ontario, respectively with 67, 67, 283, 38, 108 and 140 lines. The execution times of their algorithm with each of these instances range from 150 to 1700 seconds. In this problem, every trip component includes a constant in-vehicle travel time as well as waiting times for vehicles. They quantify the waiting time for a vehicle by using the statistical distribution of waiting times for the arrival of the first vehicle of a given transit line at a given stop. This is achieved by duplicating each node for each bus line which serves this stop. Each duplicated node is then connected to the original node by two arcs: one representing the alighting time and the other, the boarding time. Then, they assign each traveler waiting at an original node to the the first vehicle that serves any of the arcs leaving this node. For the second formulation, they propose a non-linear model, which is solved in linear time by a label-setting algorithm. However, the biggest limitation of the model is that waiting times at stops are not affected by transit volumes. Hence, every passenger may board the first arriving vehicle and queuing delays at stops because of full vehicles are not considered by the model. This formulation reduces the congestion phenomenon to a comfort problem.

Among the studies which use the model of Spiess and Florian [1989] for their lower-level problem, some differ with various features at the upper-level. For instance, Constantin and Florian [1995] propose a non-linear bi-level model with an upper-level which consist in finding the frequencies which minimize the sum of passengers path duration. A few years later, Noriega and Florian [2003] extend the model proposed by Constantin and

Florian [1995] by integrating vehicle capacities and considering congestion in the path duration of passengers. However, the resulting bi-level model is still non-linear. Yu et al. [2010] present a non-linear bi-level model with an upper level aiming to minimize total passenger travel time and with a lower level assigning passengers to paths based on the optimal strategy of Spiess and Florian [1989]. Ruisanchez et al. [2012] propose a bi-level model with an upper-level consisting of the capacity constrained assignment problem formulated by De Cea and Fernandez [1993]. They use a tabu search method to solve their bi-level model.

Martínez et al. [2014] base their work on the model of Constantin and Florian [1995], originally formulated as a nonlinear bi-level model. They use the assignment problem of Spiess and Florian [1989] as lower-level problem. The authors propose a discretization of the domain of frequencies and introduce binary variables to assign a frequency to each line. With this model, Martínez et al. [2014] obtain a linear formulation that, under certain conditions, has a single-level structure. However, the proposed method can only solve small instances. Hence, the authors develop a tabu search heuristic to solve their FSP on large-scale instances.

Few papers propose a bi-level model without using the passenger assignment of [Spiess and Florian, 1989] at their lower level. Ma and Chow [2021] propose an iterative simulation-based frequency bi-level optimization model. They use the MATSim software as the lower level to simulate passenger path choices. MATSim is an open source transit simulation tool that seeks to determine passenger choices that minimize the sum of all passengers path duration. However their model is non-convex and the solution method is a heuristic designed to obtain a local optimum. They test their method on the existing bus network of Brooklyn, New York, which is composed of 78 bus lines and carries over 650,000 passengers per weekday. In their experiment, one run requires about 11 hours.

4 Research gaps

This section focuses on enumerating the observed gaps for line planning and frequency setting problems. Some of these gaps are identified as important in view of the needs of the company Lumiplan and selected for further integration in our works.

4.1 Line Planning Problem

Concerning the line planning problem, we note a need in terms of generation of lines in a dynamic way within an algorithm. Indeed, the vast majority of papers dealing with the design of lines select lines from a line pool established before optimization. The result of line plan is therefore strongly dependent on this line pool.

Moreover, we notice the service level for passengers can be represented in several ways. In most cases, the entire demand must be met with no more than two transfers [Mumford, 2013, Bertsimas et al., 2021]. Some representations focus on separating passenger trips into two sets, one for direct trips and one for indirect trips [Borndörfer and Karbstein, 2012]. Other service level representations include measures of passenger deviation from the shortest paths in the network or from other modes [Ceder and Wilson, 1986, Baaj and Mahmassani, 1995, Gutiérrez-Jarpa et al., 2017, Bertsimas et al., 2021]. Service level representations can also include the directness of the paths available to passengers [Suman and Bolia, 2019]. Finally, other representations focus on passenger in-vehicle travel time, defined as a function of in-vehicle travel time with a penalty for each transfer [Borndörfer et al., 2009, Mumford, 2013]. We note the absence of papers separating paths by directness or non-directness, combined with the use of service level parameters. Concerning this service level aspect, in accordance with the European decree EN- and with the service level model observed in different studies, we retain a combination of two service levels: (1) A passenger must have a path to go from its origin to destination without deviating from its shortest path by a given limit. This is called maximum-spt-deviation. (2) A minimum percentage of passengers must be able to go from their origin to their destination without transfer. This is denoted minimum-rdc-percentage.

4.2 Frequency Setting Problem

Concerning the FSP, we note that the literature is particularly scarce on this topic. Indeed, the great majority of papers concentrate on the LPP, without refining the estimated frequencies. Most papers dealing with the FSP are formulated as bi-level non-linear problems and integrate the passenger assignment model of Spiess and Florian [1989] at the lower level. From then on, these problems are solved with heuristics. The bi-level formulations of Martínez et al. [2014] can be formulated as a single-level integer linear program in some cases. Nevertheless, the formulation is tractable for a solver only for small instances of the problem. We note a lack of bi-level problem formulation for the

FSP integrating operational constraints, such as bus capacity constraints, infrastructure capacity constraints or heterogeneous bus fleet constraints. We also notice that almost all studies that use a bi-level formulation maximize passenger satisfaction and that the operator perspective has been under-studied for this problem. Finally, we note that most proposed formulations are not linear and no efficient solution method has been proposed to solve realistic large scale instances.

METHODOLOGY AND CONTRIBUTIONS

This chapter first discusses the decomposition in two problems that is used in this thesis to design a bus network and determine the frequencies of these lines. In a second part, we summarize our contributions on these two optimization problems and on a case study.

1 Decomposition choices

As defined by Karbstein [2013], the Line Planning Problem (LPP) consists of finding a set of lines in an infrastructure network and their frequencies of operation such that a given travel demand can be routed. Solving the problem also implies estimating a frequency on each line. It is, in particular, necessary in order to estimate the final cost and capacity of each line in the network. Each line frequency must belong to a so called line frequency interval $[F_{min}, F_{max}]$. But as a matter of fact, this frequency is calculated with many simplifying assumptions to keep the problem tractable. Indeed, passenger demand is variable throughout the operating day. Nevertheless, the bus lines in the network remain the same throughout the day to keep it understandable and attractive for passengers. Hence, the bus network is designed for a single aggregated period of demand. This time period can be defined for a peak period or for a full day period. As a result, the selection of this time period has a major impact on the result. Other examples of simplifying assumptions that are taken in the line design process are: the formulation of the problem as a mono-level problem where passenger flows use the available capacity on the network without considering passenger choices, the use of a homogeneous bus fleet. As a result, in this thesis, we propose to solve the LPP on an aggregated period to design lines and to refine the frequencies on these lines by solving a FSP for a set of p time periods that cover a full day of demand. The sequential solving of LPP and FSP is illustrated in Figure 2.1.

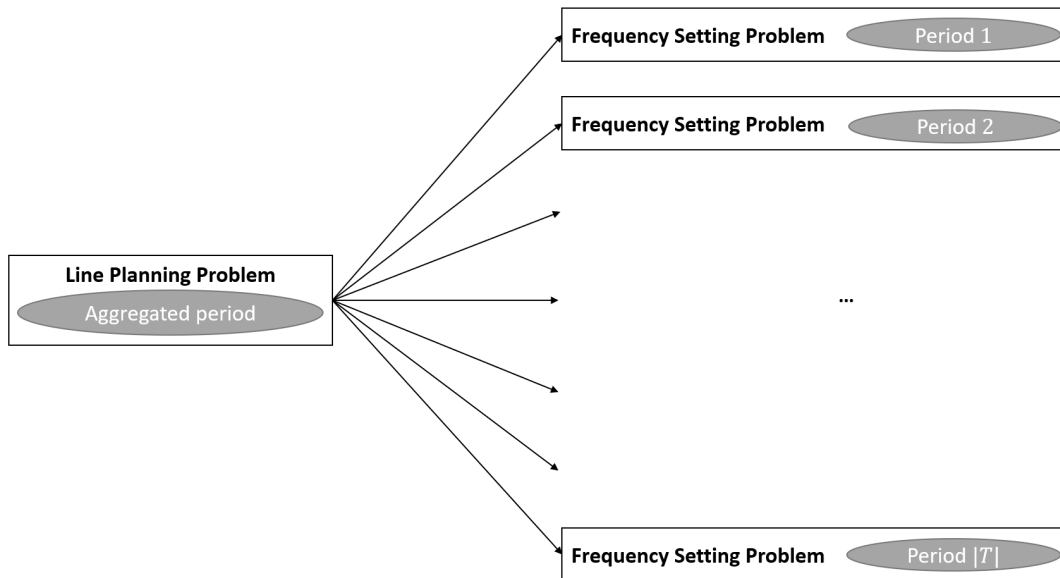


Figure 2.1 – Sequencing of the Line Planning and Frequency Setting Problems in the planning process

Once the lines are obtained, the frequencies estimated during the LPP resolution must be adapted to passenger demands that are not the same throughout the operating day. Other variations may be envisioned depending on the day of the week or the period of the year. This is where the FSP comes to define a frequency for each bus lines per direction and per time period. The FSP takes the lines defined by the LPP as input data and deals with the assignment of passengers on these lines, depending on their behaviour and for each time period. The assignment of passengers hence depends on in-vehicle travel times and expected waiting times, themselves depending on the frequency of the lines and the number of lines taken. Furthermore, the FSP integrates additional operational constraints such as a heterogeneous bus fleet or street sections capacities. As for the LPP, some simplifying assumptions are taken in the frequency setting process. Hence, the bus schedules are approximated by frequencies and the periods are solved independently. This problem solving done by period can lead to setting very different frequencies on two consecutive periods. This may not be realistic with regard to certain aspects of service continuity inherent in the activation of resources on the network.

The characteristics of the considered LPP and FSP are given in Figure 2.2.

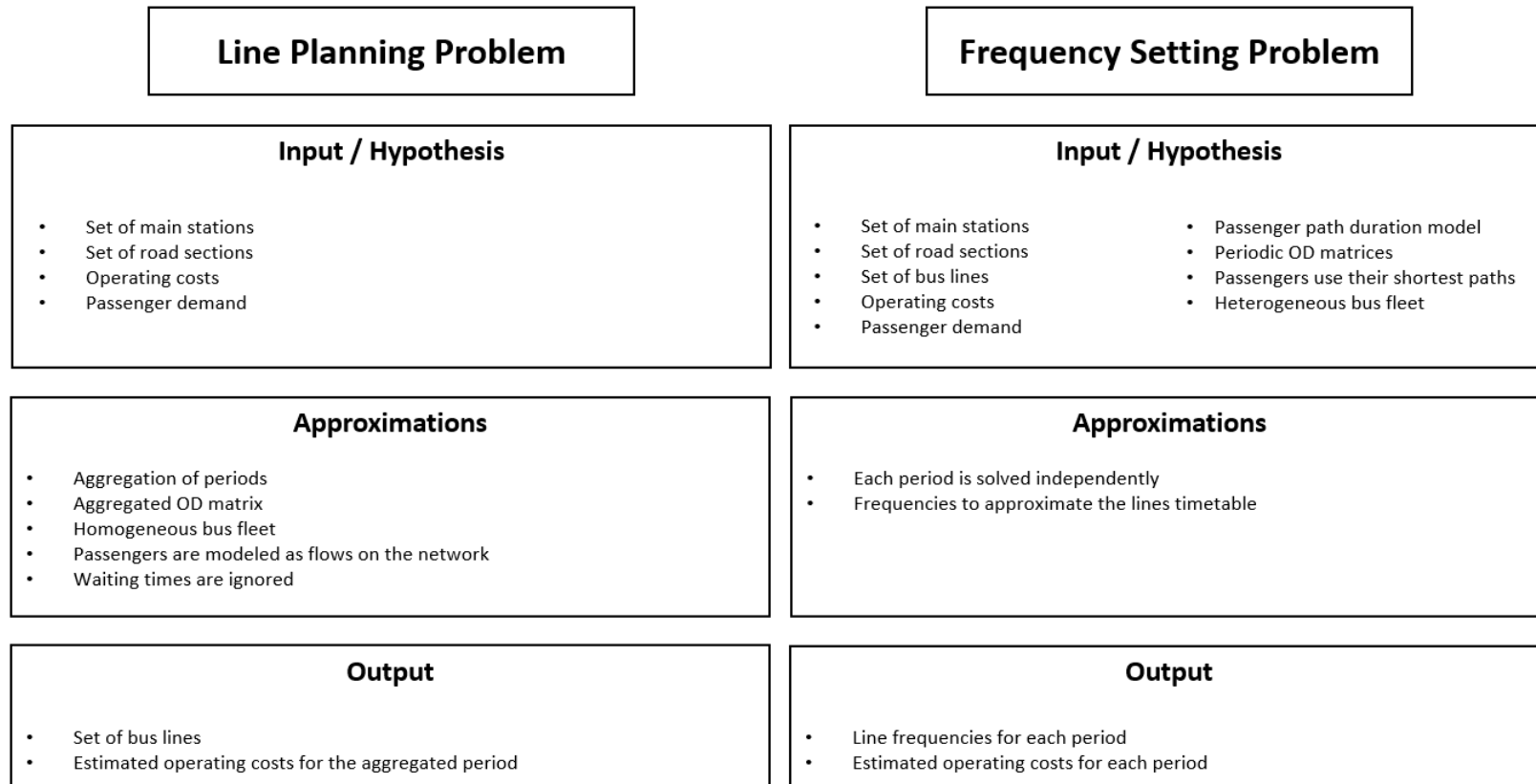


Figure 2.2 – Characteristics of the considered LPP and FSP

In addition, the timing of bus arrivals and departures is not modeled in the FSP. This synchronization is called *coordinated scheduling* [Shrivastava and O’Mahony, 2006] and allows the schedules of the different lines to be synchronized in order to optimize the connections within the hubs. Coordinated scheduling therefore reduces the waiting time for passengers during their connections. It is usually addressed in the timetable development phase of the overall public transport planning process Wu et al. [2019].

2 Summary of contributions

In this thesis, we first focus on defining the line planning and frequency setting problems according to the specific needs of Lumiplan. In addition, we propose optimization algorithms to solve each of these problems.

In Chapter 3, devoted to the LPP, we extend the DC model [Borndörfer and Karbstein, 2012] which provides a good approximation for the counting of transfers in passenger paths in a bus network design optimization model. The kilometric costs are minimized while respecting new service level parameters for passengers. The number of transfers is not limited in this model, however, a distinction is made between direct and indirect passenger paths. We integrate realistic characteristics like line frequency interval, arcs capacities and maximum number of lines, useful for public transportation companies. Moreover, to respect the service levels chosen parameters, we add some constraints to model the directness of the paths attributed to passengers. The problem is solved by an innovative process based on column generation and column enumeration methods. In particular, for our extension of the DC model, we define some labeling algorithms that are adapted to the generation of lines and paths. In our experiments, we present a comparison between our results and those of the literature on two existing instances. Our results clearly show the superiority of our integrated line generation approach over the classical a priori line pool generation.

In Chapter 4, devoted to the FSP, we investigate two major challenges, which are the minimization of the operator costs and the modeling of passengers’ behavior. To satisfy the operator’s expectations, the kilometric costs are minimized. Moreover, we assume a passenger to take its shortest path with respect to the path duration, including in-vehicle travel time and expected waiting time. More precisely, for each OD pair, the possible trips are modeled as paths. Each path is defined by a sequence of arcs, each associated with a line. The expected waiting time depends on the frequency assigned to each line.

The problem of assigning frequencies to lines is thus a bi-level problem where the leader is the bus network operator who sets the line frequencies by minimizing its costs and the followers are the passengers who minimize their path duration according to the line frequencies. We model this problem as a mixed integer linear program. The bi-level problem is formulated as a single-level problem with a quadratic number of variables and constraints depending on the number of possible paths on the network. To make this model tractable on real instances, we propose a dynamic passenger path selection process which progressively integrates passengers paths to the model.

In Chapter 5, the two algorithms for the LPP and FSP are applied to a case study based on the bus network on the agglomeration of Poitiers, France. This exposed some limitations of the models and involved practical contributions to make the output public transportation networks more realistic. For the line planning step, based on a research article [Faivre d’Arcier, 2014] and by analysis of existing networks, we derived line typology filtering rules. Moreover, we show the relevance of our process by analyzing the potential for improvement in terms of line layout. Finally, we compare the networks obtained for different service levels and show that the bus capacity parameter, which models the average bus load on the network, has a huge impact on the returned solutions. For the frequency setting step, we show the relevance of our process by analyzing the changes in terms of line frequencies compared to the existing ones. In addition, we analyze the impact of the main input parameters on the solution.

THE LINE PLANNING PROBLEM WITH SERVICE LEVEL (LPP-SL)

Contents

1	The Line Planning Problem with service level	54
1.1	Problem formulation	54
1.2	Problem modeling	56
1.3	The model	58
1.4	Classes	60
2	Column-Generation for the LPP-SL	62
2.1	Solution methodology	62
2.2	Pricing problems	64
2.3	Column enumeration step	68
3	Computational results on instances from the literature . . .	69
3.1	Comparison with literature results	69
3.2	Generating bus networks with CGEP	72
4	Conclusion	74

This section focuses on modeling the Line Planning Problem (LPP) faced by the Lumiplan company and defining an algorithm to solve this problem for medium scale networks. The LPP consists in designing the set of bus lines of a public transportation network. As shown on Figure 3.1, the LPP is located at the strategic level. Hence, decisions resulting from solving this problem falls into a long term strategic plan.

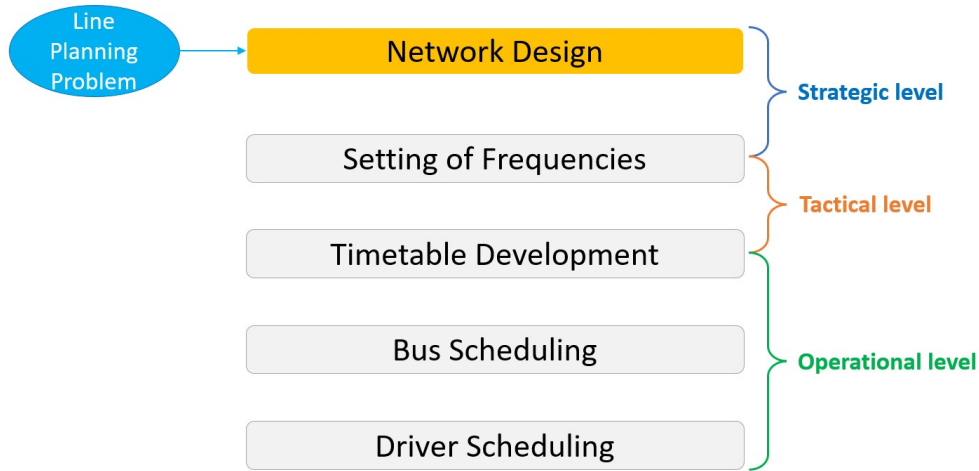


Figure 3.1 – Positioning of the Line Planning Problem in the global process of bus network planning

A literature review concerning this problem can be found on Section 2

1 The Line Planning Problem with service level

The objective is to identify the optimal set of lines and their theoretical operating frequencies, also known as the line plan, so that all passengers can travel along feasible routes while meeting operational and resource requirements, at minimal estimated cost for the operator.

1.1 Problem formulation

The infrastructure network is modeled on a graph $G = (V, A)$ where V represents a set of station nodes and A a set of arcs. Each of the arcs $a \in A$ is associated with a in-vehicle travel time r_a , a kilometric distance d_a and a maximum capacity s_a . The maximum capacity of an arc, representing the maximum number of buses allowed to run on that arc, is used to avoid saturation at the bus stations. A set of terminus stations is defined as a subset of the set of stations nodes. This set must be chosen with care. By choosing many stations, we increase the ability to generate potentially interesting lines and thus increasing the possibility of obtaining a low-cost network. However, their generation takes more time during the execution of the algorithm.

We consider passenger demands defined in an aggregated origin-destination (OD) matrix noted D . The duration for which this matrix is defined is denoted Δ . Each coefficient in this OD matrix is called an OD pair and denoted q_{st} . It represents the number of passengers wishing to travel from the station s to station t during the aggregated time period. The shortest in-vehicle travel time of an OD pair (s, t) , meaning the time of the shortest path from s to t with respect to the kilometeric distances of the arcs, is denoted by $r_{s,t}^{min}$.

To model this problem, we use equipment inputs. These inputs concern the cost per kilometer, the bus capacity, the fleet size and the maximum number of operable lines. The cost per kilometer, noted c , is an overall cost that takes into account maintenance costs, fuel costs, and driver salary costs. The idea behind this cost is that the kilometeric costs defined in the macro step of the line planning problem should be as close as possible to the actual operational costs of the operator. The bus capacity, noted κ , is the maximum number of passengers authorized in each bus. This parameter is determined by analyzing the maximum number of passengers present simultaneously on each bus currently operated and averaging these values. The fleet size, noted o refers to the number of buses simultaneously available for operation, meaning the maximum cumulative number of buses that can be operated by the operator, taking into account the availability of drivers. Finally, the maximum number of operable lines is noted n .

To consider the service level offered by a network, we retain two quality of service parameters: (1) A passenger must have a path to go from their origin to destination without deviating by a limit from the shortest path; (2) A minimum percentage of passengers must be able to go from their origin to their destination without transferring. We thus define two corresponding parameters:

- **Maximum-spt-deviation:** Defined as $\alpha \geq 1$, is the maximum deviation allowed for each passenger from its shortest path in time. It can be defined as $\frac{r_p}{r_{s,t}^{min}} \leq \alpha \quad \forall p \in \mathcal{P}_{st}$; where $p \in \mathcal{P}_{st}$ symbolizes the paths from s to t , r_p the in-vehicle travel time of the passenger path p .
- **Minimum-rdc-percentage:** Defined as $\beta \in [0, 1]$, is the minimum percentage of passengers using a rdc-passenger-path (defined below). It is based on the total number of passenger requests over the given time period associated with the OD matrix.

1.2 Problem modeling

This sub-section focuses on the presentation of the three sets of variables used to model the problem.

A bus line consists of a succession of arcs. We define a set of bus lines that we note \mathcal{L} . It is supposed that the trips of a line are the same for out- and inbound paths. The line definition can thus be restricted to a one-way trip. The one-way kilometric distance of a bus line $l \in \mathcal{L}$ is noted d_l and its in-vehicle travel time is noted r_l .

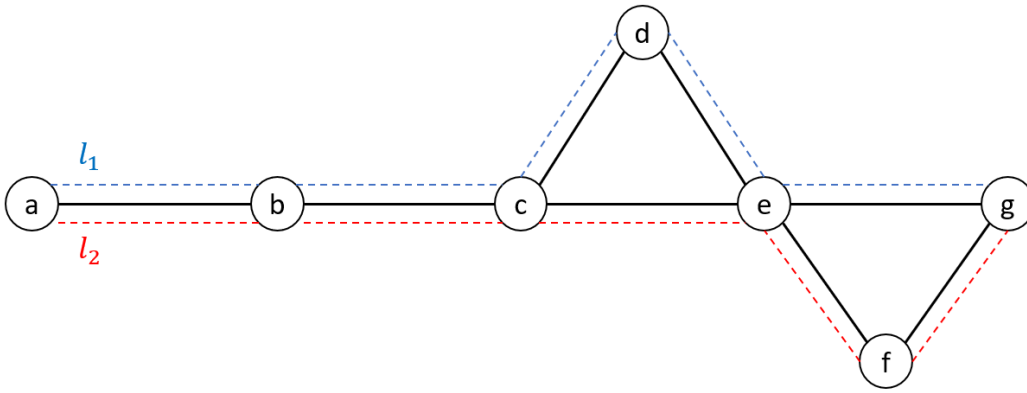


Figure 3.2 – A small network composed of seven nodes and two bus lines: $l_1 = (a, b, c, d, e, g)$ and $l_2 = (a, b, c, e, f, g)$

The notation $\mathcal{L}_{st}^\alpha(a)$ is an extension of $\mathcal{L}_{st}(a)$ firstly introduced by Borndörfer and Karbstein [2012] to represent the set of lines connecting nodes s and t through the arc $a \in A$. Hence, $\mathcal{L}_{st}^\alpha(a)$ represents the set of lines connecting nodes s and t through the arc $a \in A$ while respecting the service level parameter α .

As a bus line, a passenger path consists of a succession of arcs. We define two sets of passenger paths: rdc-passenger-paths set and indirect-passenger-paths set.

The relaxed-direct-connection (rdc) Passenger Path Set, noted \mathcal{P}^{0+} is used to approximate the set of direct passenger paths. We note \mathcal{P}_{st}^{0+} the set of rdc-passenger-paths for the (s, t) OD pair. Introduced by Borndörfer and Karbstein [2012], this set is constructed by assembling for each OD pair $(s, t) \in D$ all the existing paths in $G_{(s,t)} = (V_{(s,t)}, A_{(s,t)})$ where $V_{(s,t)}$ and $A_{(s,t)}$ respectively denote the nodes and the arcs served by at least one direct passenger path allowing to go from s to t . In other words, a passenger path is said to be a rdc-passenger-path if there exists, on each of its arcs, at least one line $l \in \mathcal{L}$ that

circulates on it and goes from the origin to the destination of the path. On Figure 3.2, the passenger path $p = (a, b, c, e, g)$ is a rdc-passenger-path from a to g but not a direct passenger path. Hence, there is no line running on the path $p = (a, b, c, e, g)$ but there is at least one direct line running from a to g on each of its arcs. The round trip kilometric distance of a rdc-passenger-path $p \in \mathcal{P}^{0+}$ is noted $d_p = \sum_{a \in p} d_a$ and its in-vehicle travel time is noted $r_p = \sum_{a \in p} r_a$.

The main interest of creating the rdc-passenger paths set is that it enables approximating the number of transfers in a column generation approach and thus reduces the complexity of pricing the path variables of connecting passengers. By accepting an over-estimation of the number of direct connections, Borndörfer and Karbstein [2012] propose a constraint, called DC (Direct Connection) constraint, to omit the specific coverage of paths by bus lines but to focus only on the arcs covered by bus lines. Thus, the use of rdc-passenger-paths instead of direct-passenger paths avoids the need to define the passenger path in terms of individual lines iterating over the lines during the pricing problems. rdc-passenger-path pricing problems are then assimilated to shortest path sub-problems. Borndörfer and Karbstein [2012] show that this approach provides a very good approximation of the actual number of direct and indirect paths in the problem solutions.

The indirect-passenger-path Set \mathcal{P}^1 contains all indirect-passenger trips, i.e., passenger trips for which at least one bus line runs on each of its arcs, but none of which can make the trip directly. On Figure 3.2, we can see for example that path (d, e, f) is indirect because there is at least one line on each of its arcs but none of these lines allows to go directly from d to f . We note \mathcal{P}_{st}^1 the set of indirect-passenger-paths for the (s, t) OD pair. The round trip kilometric distance of an indirect-passenger-path $p \in \mathcal{P}^1$ is noted $d_p = \sum_{a \in p} d_a$ and its in-vehicle travel time is noted $r_p = \sum_{a \in p} r_a$.

Table 3.1 summarizes the notations used in *LPP-SL*.

Table 3.1 – Main notation used to modelize the LPP

$G = (V, A)$	Graph representing the infrastructure
V	Set of main stations
A	Set of road sections
Δ	Time duration of the considered time period
D	OD matrix modeling passenger demand
q_{st}	Quantity of passengers requesting to go from s to t
$r_{s,t}^{min}$	Shortest in-vehicle travel time from s to t
\mathcal{P}^{0+}	Set of rdc passenger paths
\mathcal{P}_{st}^{0+}	Set of rdc passenger paths from s to t
\mathcal{P}^1	Set of indirect passenger paths
\mathcal{P}_{st}^1	Set of indirect passenger paths from s to t
\mathcal{L}	Set of bus lines
$\mathcal{L}_{st}^\alpha(a)$	Set of lines connecting s and t through arc a with a valid SPT deviation
f_{min}	Minimum frequency for an operated line
f_{max}	Maximum frequency for an operated line
n	Maximum number of lines to be operated
o	Number of operable buses
κ	Passenger capacity of buses
s_a	Arc a saturation threshold
d_a	Kilometric distance of the arc $a \in A$
r_a	In-vehicle travel time associated to the arc $a \in A$
r_l	One-way travel time of a bus on the line $l \in L$
r_p	In-vehicle travel time of the passenger path $p \in P$

1.3 The model

In this model, we use the following variables:

- $z_l = \begin{cases} 1, & \text{if the bus line } l \in \mathcal{L} \text{ is opened} \\ 0, & \text{otherwise} \end{cases}$
- $f_l \in \mathbb{N}$: operating frequency of the bus line $l \in \mathcal{L}$. It defines the number of bus passages for each of the two directions of the line during a given period. In this study, we assume line frequencies to be symmetric in each of the two directions and to be between f_{min} and f_{max} .

- $y_p^{0+} \in \mathbb{R}_+$: quantity of passengers assigned to the rdc-passenger-path $p \in \mathcal{P}^{0+}$
- $y_p^1 \in \mathbb{R}_+$: quantity of passengers assigned to the indirect-passenger-path $p \in \mathcal{P}^1$

LPP-SL

$$\min \quad \lambda \times \sum_{l \in \mathcal{L}} f_l \times c \times d_l + (1 - \lambda) \times \left(\sum_{p \in \mathcal{P}^{0+}} y_p^{0+} \times r_p + \sum_{p \in \mathcal{P}^1} y_p^1 \times r_p \right) \quad (3.1)$$

$$\text{s.t.} \quad \sum_{p \in \mathcal{P}_{st}^{0+}} y_p^{0+} + \sum_{p \in \mathcal{P}_{st}^1} y_p^1 = q_{st} \quad \forall (s, t) \in D \quad (3.2)$$

$$\sum_{p \in \mathcal{P}^{0+} \mid a \in p} y_p^{0+} + \sum_{p \in \mathcal{P}^1 \mid a \in p} y_p^1 \leq \sum_{l \in \mathcal{L} \mid a \in l} \kappa \times f_l \quad \forall a \in A \quad (3.3)$$

$$\sum_{(u,v) \in D \mid \mathcal{L}_{uv}^\alpha(a) \subseteq \mathcal{L}_{st}^\alpha(a)} \sum_{p \in \mathcal{P}_{uv}^{0+}(a)} y_p^{0+} \leq \sum_{l \in \mathcal{L}_{st}^\alpha(a)} \kappa \times f_l \quad \forall a \in A, \forall (s, t) \in D \quad (3.4)$$

$$\sum_{l \mid a \in l} f_l \leq s_a \quad \forall a \in A \quad (3.5)$$

$$\sum_{p \in \mathcal{P}^{0+}} y_p^{0+} \geq \beta \times \sum_{(s,t) \in D} q_{st} \quad (3.6)$$

$$\sum_{l \in \mathcal{L}} 2 \times r_l \times f_l \leq o \times \Delta \quad (3.7)$$

$$f_l \geq f_{min} \times z_l \quad \forall l \in \mathcal{L} \quad (3.8)$$

$$f_l \leq f_{max} \times z_l \quad \forall l \in \mathcal{L} \quad (3.9)$$

$$\sum_{l \in \mathcal{L}} z_l \leq n \quad (3.10)$$

$$y_p^{0+} \in \mathbb{R}_+ \quad \forall p \in \mathcal{P}^{0+}$$

$$y_p^1 \in \mathbb{R}_+ \quad \forall p \in \mathcal{P}^1$$

$$f_l \in \mathbb{N} \quad \forall l \in \mathcal{L}$$

$$z_l \in \{0, 1\} \quad \forall l \in \mathcal{L}$$

The *objective function* (3.1) minimizes the weighted sum of the total kilometric costs and total passenger in-vehicle travel times. A parameter $\lambda \in [0, 1]$ is used to model the weight of each of these two objectives. In this function, the kilometric cost of line is defined by the total distance traveled by buses on the line multiplied by a cost per distance specified by the operator. Total passenger in-vehicle travel times are defined by

the sum of the in-vehicle travel times of the passenger paths multiplied by the number of passengers using these paths.

Constraints (3.2) force each OD pair to be satisfied. Constraints (3.3) ensure a sufficient line capacity on each arc to transport all passengers. Constraints (3.4) ensure an approximate sufficient line capacity on each arc to transport passengers using a rdc-passenger-path. This constraint eliminates the assignment of rdc-passenger-paths to particular lines and thus the need to enumerate all bus lines during rdc-passenger pricing. Constraints (3.5) ensure the number of buses using an arc does not exceed the maximum capacity of the arc. Constraint (3.6) enforces the ratio of passengers having a rdc-passenger-path to be bigger than β . Constraint (3.7) ensures the availability of the number of buses required for the frequency selected based on the round-trip travel time of the bus lines operated. Constraints (3.8) and (3.9) are used to enforce the frequency of each line to belong to a given range. Constraint (3.10) limits the number of open lines to a maximum threshold.

Constraints (3.4) [Borndörfer and Karbstein, 2012] set a sufficiently high frequency for each of the lines on each arc to meet passenger demand using a rdc-passenger-path. They are built, for each arc $a \in A$, according to the dominance of couple of OD pairs, meaning if the line set available for the first OD pair is a subset of the line set available for the other one. The big advantage of these constraints is the omission of specific passenger path coverage by bus lines but the focus only on the arcs covered by bus lines. However, even if these DC constraints present a strong interest, they present a disadvantage: the number of direct travelers is overestimated by using them. See Appendix A.1 for an example with the number of direct travelers being overestimated.

Compared to Borndörfer and Karbstein [2012], we remove the penalty on indirect-passenger-paths from the objective function, replacing it with our service level parameter β . This parameter provides precise and direct way to favor direct passenger paths. We fix almost all the weight on the total estimated kilometric costs, but keep a small percentage on the total passengers in-vehicle travel time. Note that the overall passenger in-vehicle travel time is integrated at this step for comparison with the literature but that its weight is set to be negligible in our case study.

1.4 Classes

Equivalence classes were introduced by Borndörfer and Karbstein [2012] to reduce the complexity of the DC constraint. Equivalence classes are useful for decoupling passenger

paths from bus lines in DC constraints. According to Borndörfer and Karbstein [2012], for a graph $G = (V, A)$ where V represents the main stations and A the arcs, equivalence classes are defined according to the existing rdc-passenger-paths $p \in \mathcal{P}^{0+}$ and bus lines $l \in \mathcal{L}$.

dc-equivalence classes

For each arc $a \in A$, two OD-pairs (u, v) and (s, t) are said dc-equivalent (direct connection) with respect to a , if $\mathcal{L}_{uv}(a) = \mathcal{L}_{st}(a)$ (these OD have the same set of direct lines that include arc a). Then, (s, t) and (u, v) are said to belong to the same equivalence class for arc a . The set of equivalence classes for dc-equivalent OD-pairs with respect to a is denoted by $\tilde{D}(a) = \{[s, t]_a\}$, where (s, t) is one of the OD pairs chosen arbitrarily among all the OD pairs of the class. We note $(u, v) \in [s, t]_a$ if the two OD-pairs (u, v) and (s, t) are said dc-equivalent with respect to a .

dc-dominated classes

Furthermore, based on the equivalence classes definition, an OD pair (u, v) is said dc-dominated with respect to arc a by an OD pair (s, t) if $\mathcal{L}_{uv}(a) \subseteq \mathcal{L}_{st}(a)$. We note $(u, v) \in [s, t]_a^{\leq}$ if the OD-pair (u, v) is dc-dominated by the OD pair (s, t) with respect to a .

dc-dominated classes wrt maximum-spt-deviation

We finally introduce a definition based on the dc-dominated classes introduced by Karbstein [2013]. An OD pair (u, v) is said dc-dominated with respect to arc a and with respect to the service-level parameter α by an OD pair (s, t) if $\mathcal{L}_{uv}^\alpha(a) \subseteq \mathcal{L}_{st}^\alpha(a)$. $\mathcal{L}_{st}^\alpha(a)$ symbolizing the set of lines relying s and t through a in a travel time at most equal to $\alpha \times r_{st}^{min}$. We note $(u, v) \in [s, t]_a^{\leq, \alpha}$ if the OD-pair (u, v) is dc-dominated by the OD pair (s, t) with respect to a and with respect to the service-level parameter α .

Considering the LPP-SL model, for each equivalence class $[s, t]_a$, one constraint (3.4) is defined for each OD pair in the class. Using the dc-domination principle, constraints (3.4) can be reformulated as follows [Borndörfer and Karbstein, 2012]:

$$\sum_{(u,v) \in [s,t]_a^{\leq, \alpha}} \sum_{p \in \mathcal{P}_{uv}^{0+}(a)} y_p^{0+} \leq \sum_{l \in \mathcal{L}_{st}^\alpha(a)} \kappa \times f_l, \quad \forall a \in A, \forall [s, t]_a \in \tilde{D}(a), \quad (3.11)$$

Example 1.1. Let us illustrate the concept of equivalence classes on the example of Figure 3.2. Assuming the service level α is verified by all paths on this network, we observe that lines l_1 and l_2 connect a to c and a to g through the arc (b, c) . We thus have $\mathcal{L}_{ac}^\alpha(b, c) = \{l_1, l_2\}$ and $\mathcal{L}_{ag}^\alpha(b, c) = \{l_1, l_2\}$. As $\mathcal{L}_{ac}^\alpha(b, c) = \mathcal{L}_{ag}^\alpha(b, c)$, we thus can say that

OD-pairs (a, c) and (a, g) are dc-equivalent with respect to arc (b, c) and we denote this equivalence class by $[a, c]_{(b,c)}$.

Supposing equivalence classes are not used and defining constraints (3.4) for both OD, the following two constraints are derived:

$$\begin{aligned} - & \sum_{(u,v) \in D \mid \mathcal{L}_{uv}^\alpha(b,c) \subseteq \mathcal{L}_{ac}^\alpha(b,c)} \sum_{p \in \mathcal{P}_{uv}^{0+}(b,c)} y_p^{0+} \leq \sum_{l \in \mathcal{L}_{ac}^\alpha(b,c)} \kappa \times f_l \\ - & \sum_{(u,v) \in D \mid \mathcal{L}_{uv}^\alpha(b,c) \subseteq \mathcal{L}_{ag}^\alpha(b,c)} \sum_{p \in \mathcal{P}_{uv}^{0+}(b,c)} y_p^{0+} \leq \sum_{l \in \mathcal{L}_{ag}^\alpha(b,c)} \kappa \times f_l \end{aligned}$$

Since $\mathcal{L}_{ac}^\alpha(b, c)$ and $\mathcal{L}_{ag}^\alpha(b, c)$ refer to the same set of lines, these two constraints are identical. Hence, equivalence classes avoid duplicating these constraints.

2 Column-Generation for the LPP-SL

The LPP-SL model supposes that all feasible bus lines, direct and indirect passenger paths, are enumerated before solving the problem. This assumption reduces the direct application of the model to transportation networks of limited sizes. Accordingly, we propose a column generation approach to generate the associated variables while solving the problem.

The proposed method is composed of a first phase for column generation followed by a second phase for column enumeration. The general process is described in Section 2.1. The dual problem of the master problem is then defined with three pricing problems for the generation of bus lines and passenger paths. The enumeration step is then described in Section 2.3.

It should be noted that, because the lines and paths are generated independently in this approach, while their variables are in the same constraints, the resulting solution is not proven optimal.

2.1 Solution methodology

Our general solution methodology is outlined in Figure 3.3. This method is based on the approach of Baldacci et al. [2008]. Its general principle has been applied by Glize et al. [2020] to solve multi-vehicle covering tour problems. We denote it *Column Generation and Enumeration Process (CGEP)*. The first step is called the column generation step (CG). Some initial passenger paths are generated based on the existing lines of the network. In the case of a new network, a module based on the so called *route generation algorithm*

of Ceder and Wilson [1986] generates bus lines until we obtain a set of lines satisfying the service level parameters chosen by the user. Then, a set of initial passenger paths is generated based on these bus lines. Hence, for each generated bus line, all direct and one-transfer passenger paths are added to an initial set of passenger paths. For OD pairs for which no direct nor one-transfer path are generated for this initial line set, some two-transfers passenger paths are generated and added to the initial passenger paths set. This gives rise to a restricted formulation of the LPP-SL (denoted LPP-SL). The LPP-SL is linearized to model the use of a line and its frequency as a single variable. The linearization of the LPP-SL is described in Appendix A.2. This forms the column generation master problem. Then, bus lines, rdc-passenger-paths and indirect-passenger-paths are generated in the column generation step with two dedicated labeling algorithms: one each for bus lines and passenger paths. This process terminates when the optimum of the master problem is found, providing a lower bound to the LPP-SL. All generated columns are then used in the LPP-SL to solve the MIP with a commercial solver, thus providing an upper bound to the problem. A column enumeration (CE) step is launched with upper and lower bounds determined in order to reinforce the set of bus lines and passenger paths generated. All generated and enumerated columns are then used in the LPP-SL which is solved with a commercial solver.

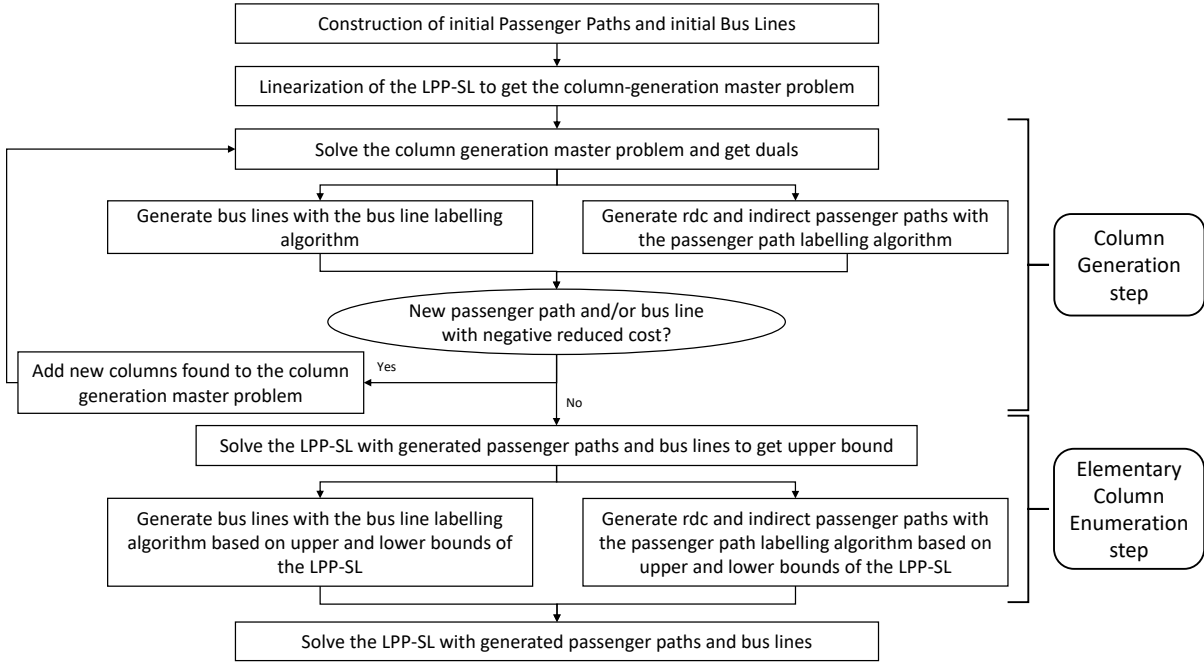


Figure 3.3 – General overview of the column generation and enumeration process (CGEP) implemented to solve the LPP-SL

2.2 Pricing problems

We propose two labeling algorithm to generate bus lines and passenger paths for our line planning problem with service level. These two labeling algorithms both present new contributions: (i) This is the first time a bus line generation algorithm is proposed for the LPP with the DC constraint to model transfers. (ii) For the bus line labeling algorithm, the storage of passengers potentially interested by a bus line label is introduced and an update of the reduced cost of the line label is performed each time a passenger can use the line in a direct way. (iii) For the passenger path labeling algorithm, a verification that the passenger path label can be continued within the service level parameter α is performed each time a new node is added to the label.

2.2.1 Bus lines

The bus line pricing problem is used to generate bus lines to integrate into the bus line set \mathcal{L} . Each of these bus lines has an associated reduced cost c_l , defined by referring

to the first constraint (8) of the dual problem (detailed in A.3).

$$\begin{aligned}
c_l = & \lambda \times c \times d_l - \kappa \times \left(\sum_{a \in l} \mu_a + \sum_{a \in l} \sum_{[s,t]_a \in \tilde{D}(a) \mid l \in \mathcal{L}_{st}^\alpha(a)} \nu_{a,[s,t]_a} \right) \\
& + \sum_{a \in l} \eta_a + \theta \times 2 \times r_l + \omega \qquad \forall l \in \mathcal{L} \qquad (3.12)
\end{aligned}$$

Constraints (3.12) are based on the μ , ν , η , θ and ω dual variables respectively referring to the constraints (3.3), (3.11), (3.5), (3.7) and (5) of the master problem.

A bus line label ϵ is defined by the tuple (v, c, δ, N, R, S) where v denotes the last visited node, c the reduced cost of the label, δ the distance of label, N the set of nodes visited by the label, R the set of OD pairs satisfied by the label because they have a direct path on its bus line and S the set of OD pairs for which no direct path exist on the label's line, whatever its future extension. The purpose of the sets R and S is to identify which OD pairs can still have a direct path on the line for further extensions. As a reminder, each acceptable bus line must have a length between l_{min} and l_{max} . The algorithm is executed by attempting to generate lines from a node src to a node dst . The initial label is thus defined as: $E_0 = (src, 0, 0, \{src\}, \{\}, \{\})$. The structure of the bus line labeling algorithm is based on four actions: (1) selection of an incomplete label, (2) extension of this label, (3) update of passengers that can have a direct path on the label's line (4) update of passenger that can no longer have a direct path on the label's line (5) dominance update of this label. This third step performs an update of the S attribute of the label while the fourth step performs an update of the R attribute. See Appendix A.3 and Appendix A.4 for more details on master and dual problems and bus line labeling algorithm.

2.2.1.1 Label extension step The label extension step is based on extension rules that must be respected when extending a line label ϵ along arc (v_ϵ, j) to create a new label ϵ' . These rules are defined as follows:

- $v_{\epsilon'} \leftarrow j$
- $c_{\epsilon'} \leftarrow c_\epsilon + \lambda \times c \times (d_{v_\epsilon, j} + d_{j, v_\epsilon}) - (\mu_{(v_\epsilon, j)} + \mu_{(j, v_\epsilon)}) \times \kappa + \eta_{(v_\epsilon, j)} + \eta_{(j, v_\epsilon)} + \theta \times (r_{v_\epsilon, j} + r_{j, v_\epsilon})$
- $\delta_{\epsilon'} \leftarrow \delta_\epsilon + (d_{v_\epsilon, j} + d_{v_\epsilon, j})$
- $N_{\epsilon'} \leftarrow N_\epsilon \cup \{j\}$
- $\text{cost_Line_Update}(c_{\epsilon'}, S_{\epsilon'})$

— $\mathbf{R_Update}(R_{\epsilon'})$

cost_Line_Update function

$\mathbf{cost_Line_Update}$ updates the reduced cost $c_{\epsilon'}$ and the set $S_{\epsilon'}$ of a label ϵ' based on the equivalence classes defined on arc $a = (v_{\epsilon}, j)$. More precisely: for an equivalence class $[s, t]_a$ with $a = (v_{\epsilon}, j)$, if s belongs to N_{ϵ} and t is equal to j or the opposite, if the duration of the path from s to t on the created line respects the maximum-spt-deviation, the reduced cost $c_{\epsilon'}$ is updated as follows:

$$c_{\epsilon'} \leftarrow c_{\epsilon'} - \sum_{a \in A_{N_{\epsilon'}(s,t)}} (\nu_{a, [s,t]_{\bar{a}}^{\leq, \alpha}} + \nu_{\bar{a}, [s,t]_{\bar{a}}^{\leq, \alpha}}) \times \kappa$$

where $A_{N_{\epsilon'}(s,t)}$ represents all the arcs on the created line between s and t and \bar{a} represents the reversed arc of a . Finally, the OD pairs that are dc-equivalent to (s, t) with respect to a , and thus belong to the equivalence class $[s, t]_a$, are added to the set $S_{\epsilon'}$.

R_Update function

The $\mathbf{R_Update}$ function identifies OD pairs that can no longer have a direct passenger path on the label ϵ' to add them to the $R_{\epsilon'}$ set. Suppose that a label ϵ' is obtained by extending a label ϵ to a node j . For each OD pair $(s, t) \notin R_{\epsilon'} \cup S_{\epsilon'}$, the sum of the travel time on the label ϵ' from s to j and the shortest in-vehicle time from j to t is computed. If it exceeds the maximum-spt-deviation, then the OD pair can no longer have a direct passenger path on the label, regardless of future extensions to other nodes. These OD pairs are then added to the set $R_{\epsilon'}$.

2.2.1.2 Dominance check The dominance check consists in comparing the labels with each other to get rid of those that are proven not to lead to an optimum path [Marquez, 2020]. A label ϵ_1 is dominated by a label ϵ_2 if the following four conditions are met:

1. $v_{\epsilon_1} = v_{\epsilon_2}$
2. $c_{\epsilon_1} \leq c_{\epsilon_2}$
3. $\delta_{\epsilon_1} \leq \delta_{\epsilon_2}$
4. $R_{\epsilon_1} \cup S_{\epsilon_1} \subseteq R_{\epsilon_2} \cup S_{\epsilon_2}$

See Appendix A.4 for a small example on the update of the reduced cost and on the dominance step.

2.2.2 Passenger paths

The passenger path pricing problem is used to generate passenger paths to integrate into the passenger path sets. This sub-problem can be used to generate both rdc-passenger-paths and indirect-passenger-paths. For this, rdc-passenger-paths are associated to the reduced cost c_p^{0+} (3.13) while indirect-passenger-paths are associated to the reduced cost c_p^1 (3.14).

$$c_p^{0+} = -\pi_{uv} + \sum_{a \in p} \mu_a + \sum_{a \in p} \sum_{[s,t]_a \in \tilde{D}(a) \mid (u,v) \in [s,t]_a^{\leq, \alpha}} \nu_{a,[s,t]_a} + (1 - \lambda) \times r_p - \psi_1 \quad \forall (u, v) \in D, \forall p \in \mathcal{P}_{uv}^{0+} \quad (3.13)$$

$$c_p^1 = -\pi_{uv} + \sum_{a \in p} \mu_a + (1 - \lambda) \times r_p \quad \forall (u, v) \in D, \forall p \in \mathcal{P}_{uv}^1 \quad (3.14)$$

(3.13) and (3.14) are based on the π , μ , ν and ψ_1 dual variables respectively referring to the constraints (3.2), (3.3), (3.11) and (3.6) of the master problem. A passenger path label is defined by the tuple (v, c, τ, N) where v represents the last node visited by the label, c the reduced cost of the label, τ the time of the label and N the set of nodes visited by the label. Recall that every passenger path must respect the service level parameter α . The structure of the common path generation pricing problem is based on three actions : (1) selection of an incomplete label, (2) extension of this label, (3) dominance checking of this label. See Appendices A.3 and A.5 for more details on master and dual problems and on the passenger path labeling algorithm.

The **label extension step** is based on extension rules that must be respected when a path label ϵ is extended along an arc (v_ϵ, j) to create a new label ϵ' . They are defined as follows:

- $v_{\epsilon'} \leftarrow j$
- $c_{\epsilon'} \leftarrow c_\epsilon + (1 - \alpha) \times r_{v_\epsilon, j} + \mu_{(v_\epsilon, j)}$
- $\tau_{\epsilon'} \leftarrow \tau_\epsilon + r_{v_\epsilon, j}$
- $N_{\epsilon'} \leftarrow N_\epsilon \cup \{j\}$

Each time a label is extended along an arc (i, j) , it must be checked whether the label meets certain conditions. In the case of a rdc-passenger-path generation, a search is made in the bus lines pool \mathcal{L} to determine whether there is at least one line that includes the arc (i, j) and covers the origin and destination of the path with a in-vehicle travel time respecting the service level parameter α . In the case of an indirect-passenger-path, a

search is made in the bus lines pool \mathcal{L} to determine the existence of at least one line including the arc (i, j) . See Appendix A.5 for a small example on this update of the reduced cost.

Finally, a function `cost_rdcPath_Update`, dedicated to the rdc-passenger-paths generation, is executed each time the label is extended to a new node. Let src and dst be the origin and the destination of a rdc-passenger-path and let a denote the arc that extends the label ϵ to generate ϵ' . Let $[s, t]_a$ be an equivalence class that dominates the OD pair (src, dst) with respect to a . The set of all such equivalence class is $\{[s, t]_a \in \tilde{D}(a) \mid (src, dst) \in [s, t]_a^{\leq, \alpha}\}$. For each equivalence class in this set, the reduced cost of the label $c_{\epsilon'}$ is updated according to the values of the dual variables $\nu_{a, [s, t]_a}$.

$$c_{\epsilon'} \leftarrow c_{\epsilon} - \sum_{[s, t]_a \in \tilde{D}(a) \mid (src, dst) \in [s, t]_a^{\leq, \alpha}} \nu_{a, [s, t]_a} \quad (3.15)$$

Following this, a **dominance check** is performed between the extended label and the others. To do this, a label ϵ_1 is said to be dominated by a label ϵ_2 if the following four conditions are met:

- $v_{\epsilon_1} = v_{\epsilon_2}$
- $c_{\epsilon_1} \leq c_{\epsilon_2}$
- $\tau_{\epsilon_1} \leq \tau_{\epsilon_2}$
- $N_{\epsilon_1} \subseteq N_{\epsilon_2}$

2.3 Column enumeration step

Only using column generation method to get an optimal solution to the line planning problem with service level may be not sufficient. In fact, columns leading to an optimal solution of the LPP-SL may not be generated because column generation is not used at each node of the branch-and-bound algorithm that solves the MIP. To consolidate the sets of bus lines and passengers paths, a column enumeration step is added as introduced by Baldacci et al. [2008] for the Vehicle Routing Problem.

This column enumeration step is based on the enumeration of all columns whose reduced cost is less than the difference between the best known upper bound z_{IP} of the problem and its lower bound z_{MP} determined at the output of the column generation step. Indeed, each bus line and passenger path having a reduced cost smaller or equal

than $z_{IP} - z_{MP}$ is added to the problem which is then solved with the MIP solver, using the previous upper bound as an initial solution.

3 Computational results on instances from the literature

Surprisingly, the literature and publicly available benchmarks is extremely scarce for the LPP. In this section, we evaluate our CGE method on two instances from the literature. The data for these benchmarks were kindly shared by Marika Karbstein, who is duly acknowledged here. The models are implemented with Julia and solved by CPLEX 20.1 through the JuMP interface on a DELL R440 1U server with a 2.1 GHz Intel Xeon 6230 CPU and 192GB of RAM. All details concerning the instances can be found in Appendix C.

3.1 Comparison with literature results

To evaluate the efficiency of our approach, we take Borndörfer and Karbstein [2012] as a reference, since these authors also use the DC relaxation. The main difference between the two approaches is the generation of bus lines performed in CGEP, compared to the a priori enumeration of a line pool of Borndörfer and Karbstein [2012].

To be able to compare our results to those of Borndörfer and Karbstein [2012], our model is adapted to solve the same problem. For both instances, all nodes are set as potential termini. Furthermore, operational constraints concerning the maximum number of lines and the maximum bus capacity are relaxed. In the same way, the two service level constraints concerning parameter α and β are both relaxed. In addition, the definition set of line frequencies is modified so that it can be defined in the same way as Borndörfer and Karbstein [2012]. Thus, line frequencies can be set to 3, 6, 9 or 18. Furthermore, the objective function and thus the bus line generation sub-problem are both modified by adding a fixed opening cost to each bus line. For its part, as mentioned before, the overall passenger in-vehicle travel time is integrated in the objective function for comparison with the literature. With these modifications, our model becomes comparable to Borndörfer and Karbstein [2012].

The parameters presented below are those used by Borndörfer and Karbstein [2012] for their benchmarks. They are summarized in Table 3.2.

Table 3.2 – Borndörfer and Karbstein [2012] parameters for Sioux-Falls and Dutch. The columns list the weight λ in the objective function, the kilometric cost C , the buses maximum capacity K , the transfer penalty σ (in minutes) for each indirect-passenger-path used.

Instance	Weight λ	Km. Cost	Fixed Cost	Max Bus Load	Transfer Penalty
Dutch	0.8	100	100	600	15
Sioux-Falls	0.8	1.96	100	57	15

For both instance Dutch or Sioux-Falls, Borndörfer and Karbstein [2012] base their experiments on three cases: DC1, DC2 and DC3. These cases differ from each other in the number of lines present in the upstream generated line pool. For each of these cases, the size of the line pool is presented in Tables 3.3 and 3.4 in order to compare the number of lines considered by Borndörfer and Karbstein [2012] with the number of lines generated by our program. For our experiment, we use our process, first using the CG step alone to generate bus lines and passenger paths then using the complete CGEP.

3.1.1 Dutch instance

Table 3.3 – Statistics on the Dutch instance for the first experiment. The columns list the number of lines considered, the objective value of the LPP-SL, the in-vehicle travel times (in minutes) of passengers, the kilometric costs, the number of direct passengers expected by the model and the solving time. The CG and CGEP rows present our approach with CG step alone or with CGEP.

Method	# lines considered	Obj value	Travel times	Kilometric costs	Dir. Pass predicted	Solving time
DC1	402	2613305	12790925	68900	179496	15 sec
DC2	2679	2612122	12790610	67500	180644	1h
DC3	7302	2614374	12811870	65000	179384	10h (tl)
CG	349	2619250	12797448	74700	178720	28 sec
CGEP	2889	2608912	12777358	66800	180138	97 sec

Table 3.3 summarizes our results compared to the results of Karbstein [2013] on the Dutch instance; see Appendix C for more details on this instance. For this experiment, we notice that the CE step is necessary to obtain a better solution than Karbstein [2013]. Indeed, by not using this step, we obtain a solution 0.27% worse than the best of their solutions, while by using CE we obtain a solution 0.12% better. This 0.12% better solution

is obtained in less than two minutes with a dynamically generated pool of 2889 lines while the best solution of Karbstein [2013] is obtained in one hour with an a priori generated line pool composed of 2679 lines. Note that CPLEX solves all MIPs of Karbstein [2013] to optimality except for DC3 where the final solution is returned with an integrality gap of 0.05%.

3.1.2 Sioux-Falls instance

Table 3.4 – Statistics on the Sioux-Falls instance for the first experiment. The columns list the number of lines considered, the objective value of the LPP-SL, the in-vehicle travel times (in minutes) of passengers, the kilometric costs, the number of direct passengers predicted by the model, the solving time and the gap to optimality. The DC variants are those presented in Borndörfer and Karbstein [2012]. The CG and CGEP lines present our approach with the single CG step or with CGEP.

Method	# Lines Considered	Obj Value	Travel Times	Kilometric Costs	Dir. Pass Predicted	Resolution Time
DC1	866	666011	3294471	8896	360600	10h
DC2	9397	687911	3417999	5389	360600	10h
DC3	15365	675840	3361196	4501	360400	10h
CG	450	638648	3176048	4298	360600	20 min
CGEP	26861	638648	3176048	4298	360600	5h

Table 3.4 summarizes our results compared to the results of Karbstein [2013] on the Sioux-Falls instance; see Appendix C for more details on this instance. We see that the column enumeration (CE) step has no effect on the obtained solution. Indeed, a solution 4.11% better than Borndörfer and Karbstein [2012] is obtained in 20 minutes without CE, while the solution of Borndörfer and Karbstein [2012] takes 10 hours. Our solution is obtained with only 450 lines generated during the column generation process. This number is quite smaller than the number of lines considered by Borndörfer and Karbstein [2012] in their sub instances DC1, DC2 and DC3. Hence, CG generates high quality bus lines for this network.

Concerning the travel time of the passengers, CGEP provides a solution that is 3.58% better than the best result of Borndörfer and Karbstein [2012]. Regarding the kilometric cost, CGEP improves the result of Borndörfer and Karbstein [2012] by 4.51%. In this case, the optimal solution of CGEP is obtained after 20 minutes at the end of the CG step. The direct passengers predicted by the model are the same as those predicted by Borndörfer and Karbstein [2012] with their sub instances DC1 and DC2.

3.2 Generating bus networks with CGEP

For this thesis, we generate two networks through our process, using our parameters and objective function. The goal is to be able to reuse these two networks in the frequency setting part. As for the previous experiments, we use the CG step alone to generate bus lines and passenger paths and we compare the results and solving times to using the CGEP approach restricted to the CG step finally using the complete CGEP. The problem parameters are summarized in Tables 3.5 and 3.6. Some of these characteristics are taken from Borndörfer and Karbstein [2012], the others are adapted to better represent Lumiplan’s view on the problem.

Table 3.5 – The columns list the weight λ in the objective function, the kilometric cost C , the buses maximum capacity K , the transfer penalty for each indirect passenger path.

Instance	Weight λ	Km. Cost	Fixed Cost	Max Bus Load	Transfer Penalty
Dutch	0.9999	100	0	600	0
Sioux-Falls	0.9999999	1.96	0	600	0

Table 3.6 – The columns list the maximum number of lines to operate, the range of possible frequencies for lines and the range of possible lengths for lines.

Instance	Time period	Max # lines	Frequency range	Length range
Dutch	120 mns	6	[1, 24]	[0, 25]
Sioux-Falls	120 mns	7	[1, 24]	[0, 0.1]

Futhermore, we set our maximum-spt-deviation parameter α to 1.5 and the minimum-rdc-percentage parameter β to 80%.

3.2.1 Dutch instance

Table 3.7 – The columns list the number of generated lines, rdc-passenger paths and indirect passenger paths, the objective value, the solving time and the gap to the best upper bound estimated by CPLEX when the solver time-limit is reached.

Method	# lines	# rdc paths	# ind. paths	Obj value	Km. costs	Solving time	CPLEX MIP gap
CG	314	320	778	68112	66800	33s	0%
CGEP	3887	963	1227	62082	60800	10h (tl)	19.85%

Index	Bus line	LPP frequency
Line n°1	[1, 21, 10, 9, 18, 17]	11
Line n°2	[2, 21, 4, 13]	7
Line n°3	[4, 19, 9, 18, 6, 21, 7, 20, 15]	7
Line n°4	[7, 21, 3, 19, 10]	12
Line n°5	[8, 5, 22, 21, 18, 9, 19, 3, 2, 11, 16]	7
Line n°6	[14, 12, 22, 13, 3, 19, 4, 21, 1, 23]	6

Table 3.8 – Lines generated by CGEP when solving the LPP-SL on the Dutch instance

For the Dutch instance, we notice that the CGEP solving times vary drastically by increasing the weight of the operating costs in the objective function. Indeed, by only increasing the weight parameter in the objective function from 80% to 99.99%, the complete CGEP solving times increases from 97 seconds with optimality proof to over 10 hours with a gap of 19.85% after reaching the time limit.

3.2.2 Sioux-Fall instance

Table 3.9 – The columns list the number of lines, generated rdc-passenger paths and indirect passenger paths, the objective value, the Solving time and the gap to the best upper bound estimated by CPLEX when the solver time-limit is reached.

Method	# lines	# rdc paths	# ind. paths	Obj value	Km. costs	Solving time	CPLEX MIP gap
CG	220	682	1539	21.0042	20.6662	59s	0%
CGEP	3160	2119	2614	19.3323	18.9924	10h (tl)	12.12%

Index	Bus line	LPP frequency
Line n°1	[2, 6, 8, 16, 17, 19, 15, 22, 23, 24, 13]	15
Line n°2	[2, 1, 3, 12, 11, 10, 16]	15
Line n°3	[4, 5, 9, 10, 11, 14, 23, 24, 13, 12]	14
Line n°4	[4, 5, 6, 8, 7, 18, 16, 10, 15, 22, 21]	21
Line n°5	[7, 18, 20, 19, 15, 14, 11, 4, 3]	12
Line n°6	[12, 13, 24, 21, 22, 20, 19, 17, 16, 10, 9]	14
Line n°7	[22, 23, 14, 11, 10, 15, 19, 17, 16]	8

Table 3.10 – Lines generated by CGEP for when solving the LPP-SL on the Sioux-Fall instance

For the Sioux-Falls instance, we notice that the CG and CGEP solving times vary drastically by increasing the weight of the operating costs in the objective function. Indeed, by increasing the weight parameter in the objective function from 80% to 99.9999% and increasing the bus capacity from 57 to 600, solving with CGEP without CE decreases from 20 minutes to 1 minute, with both MIP solved to optimality. Conversely, the complete CGEP solving time increasing from 5h for an optimal solution of the MIP to over 10 hours with an estimated CPLEX gap of 12.12%.

4 Conclusion

In this chapter we introduced a formulation of the LPP that meets the requirements of Lumiplan. This resulted on the definition of the LPP-SL where a cost oriented objective functions is minimized while service level constraints on transfers and traveled distances are satisfied. To model this problem, we selected a model from the literature where the number of transfers is approximated by an evaluation of the number of direct passenger paths. This model is extended to integrate our objective and constraints.

To solve real size instances of the LPP-SL, we presented a general methodology. It is composed of column generation and column enumeration steps to minimize the kilometric costs of a typical operational day, the value of which depends on the selected lines and their associated daily frequencies. We extend related approaches from the literature by proposing a column generation procedure that generates bus lines. Computational results for two instances from the literature show the efficiency of the approach compared to the work that inspired the model. Indeed, generating bus lines within our CGEP provides

better solutions than selecting them from a previously generated pool.

The application of this model and method to the case study that has been developed in this thesis is presented on Chapter 5.

THE FREQUENCY SETTING PROBLEM (FSP)

Contents

1	The Frequency Setting Problem	78
1.1	Problem formulation	78
1.2	The model	84
2	Solving process	87
2.1	Generating a set of passenger paths	87
2.2	Passenger Path Selection Process	89
3	Experimentation using instances from the literature	91
3.1	Experimentation settings	91
3.2	Application of PPSP	92
3.3	Frequencies Details for Dutch network	94
3.4	Frequencies details for Sioux-Falls network	94
4	Conclusion	95

This chapter focuses on the Frequency Setting Problem (FSP). In the previous chapter, we discussed the Line Planning Problem (LPP) that provides preliminary frequencies. These frequencies are necessary in the LPP to estimate the cost and capacity of the network while designing the lines. However, these frequencies need to be refined to better integrate aspects related to the variation of demand along the day, asymmetric traffic, passenger choices of path according to their travel duration and other aspects such as the different bus types that can be dispatched on the network. Solving the FSP allows this adjustment to be made by assigning frequencies to the lines of a transportation network for a different set of time periods and depending on the operated direction on the line.

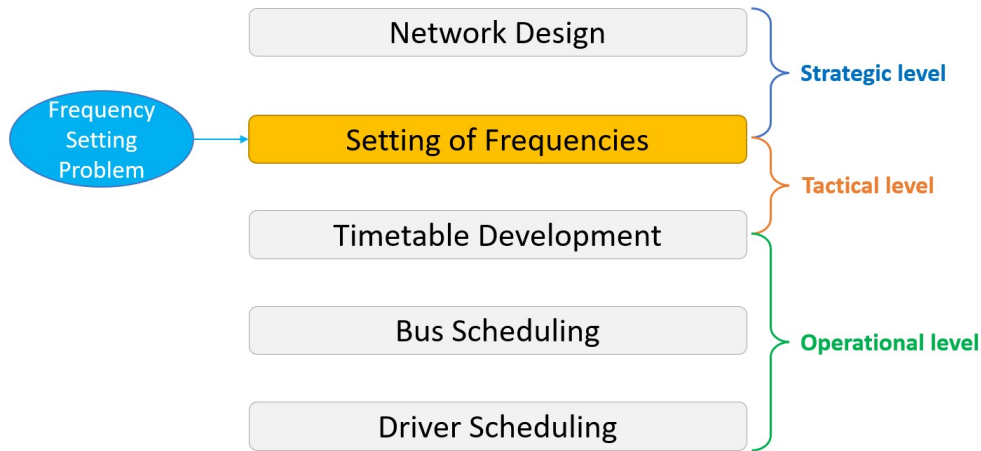


Figure 4.1 – Positioning of the Frequency Setting Problem in the global process of bus network planning

As drawn on Figure 4.1, the FSP is located at both strategic and tactical levels. It can be executed during the redesign of the network or seasonally in order to adapt the line frequencies to the passengers demands which may change over time. In addition, we integrate service level criteria and shortest path constraints, assuming passengers are taking one the paths of minimal duration. This duration is modeled more accurately by integrating the in-vehicle travel time of the travellers and their waiting times. A literature review on this problem has been presented in Section 3.

1 The Frequency Setting Problem

The FSP is formulated in the first part of this section, introducing the main related notation. In the second part, we provide a mixed integer linear program that models the studied FSP.

1.1 Problem formulation

We presume that the transportation network is modeled by a graph $G = (V, A, \mathcal{L})$. Let us consider a set of bus lines denoted \mathcal{L} . Those lines circulate on sequences of street sections which are represented by the arcs of the set A . These arcs connect the main stations of the network, which are represented by the set V . We assume that street sections are oriented and connect bus stops. Furthermore, each street section $a \in A$

is associated to a distance, an in-vehicle travel time and a saturation threshold. The saturation threshold s_a of a street section a represents the maximum number of buses that can circulate on a street section for a given period. In our study, we assume, as a first approximation, that a frequency can be assigned to each line for each period, independently of the other periods.

Hence, for a given time period, we consider an origin-destination matrix, denoted D containing the passengers' demand during the period. The duration of the time period is noted Δ . We denote by q_{uv} the estimated number of passengers having a demand from the main station u to the main station v during the time period. Users demands at intermediate stations are aggregated at the main stations.

The operating frequencies that the bus lines can be assigned to are contained in a discrete set, noted Θ_f . Θ_f is defined for $f \in \mathcal{F} = [1, |\mathcal{F}|]$. The one-way travel time of a bus on a line l during the period considered is noted r_l . Each bus line can be associated with different types of buses. The set of bus types is denoted by \mathcal{B} . Each type $b \in B$ is associated with a passenger capacity which is noted κ_b . The maximum number of available buses of type b is noted o_b .

To travel on the network, passengers use passenger paths defined in a set \mathcal{P} . Each passenger path is associated with an OD pair. A passenger path is described as a succession of arcs that connect the origin to the destination of the OD pair. Each arc in a passenger path must be associated with a bus line. A passenger path p , associated to the line l , going from node i to node k and passing through node j is defined as $p = [((i, j), l), ((j, k), l)]$. Hence, if two bus lines are serving the same street section, two passenger paths are considered on this street section. We define the *path duration* t_p of a path p as the sum of its *in-vehicle travel time* r_p and its *expected waiting time* w_p . w_p is the sum of waiting times that happen on boarding at the first stop of the path and on each transfer on the path. We denote by \mathcal{P}_{uv} the subset of feasible passenger paths for an OD pair $(u, v) \in V^2$. In this section, a path p is considered shorter than a path p' if its path duration is less than or equal to the path duration of p' .

If two consecutive arcs on a passenger path correspond to different lines, this means that a transfer occurs at the station that connects the two arcs. In our definition of the FSP, we allow for a maximum number of two transfers in a passenger path.

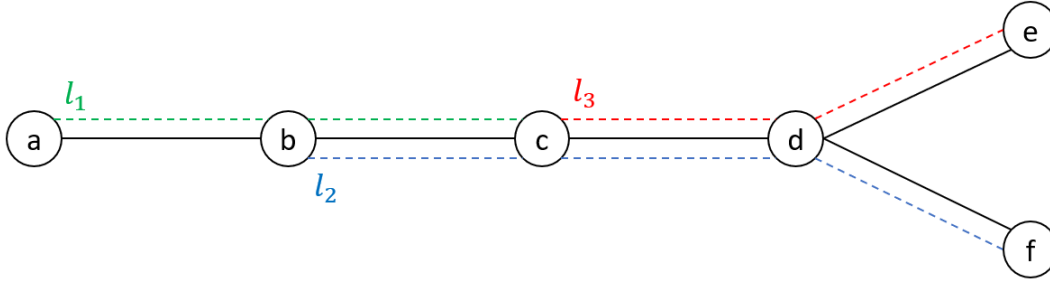


Figure 4.2 – Example of passenger paths on a small network composed of 6 nodes and with 3 bus lines circulating on it : $l_1 = (a, b, c)$, $l_2 = (b, c, d, f)$, $l_3 = (c, d, e)$

These definitions are illustrated on the example of Figure 4.2. In this example, an OD pair (b, d) is associated with a strictly positive demand. We have three possible passenger paths for this OD pair. These passenger paths are defined as $p_1 = [((b, c), l_1), ((c, d), l_2)]$, $p_2 = [((b, c), l_1), ((c, d), l_3)]$, $p_3 = [((b, c), l_2), ((c, d), l_2)]$.

Operator expenses are expressed as the sum of the kilometric costs of each operated line. The kilometric cost of a line $l \in \mathcal{L}$ is defined, for each type of bus $b \in \mathcal{B}$, as the product of its frequency and an operational cost $c_{l,b}$. The operational cost depends on the kilometric distance of the line and a fixed cost per kilometer, chosen by the operator. This fixed cost includes fuel costs as well as obsolescence costs related to the anticipation of a future buy-out when the bus retires.

The line frequencies impact both the operator cost, the capacity of the network and the service level offered to passengers. A key service level indicator is the duration of the shortest paths that will be offered to the passengers. The estimation of waiting times on these paths according to the selected line frequencies is necessary to accurately model paths duration.

Let us first consider the modeling of waiting time for the first bus on a passenger path. A common approach is, for the first bus on a passenger path, to take the average waiting time assuming uniform passenger arrival between two buses. We thus consider the average time between the passage of two buses over the period, divided by two. In case of one or two transfers, by representing this on a figure, we can clearly observe that it is not linear. The waiting time function can nevertheless be represented as a piecewise-linear 4.3 or a discrete function. Figure 4.3 is obtained for a time period of 60 minutes and possible line frequencies ranging from 1 to 12. It is not too difficult to model this piece-wise linear

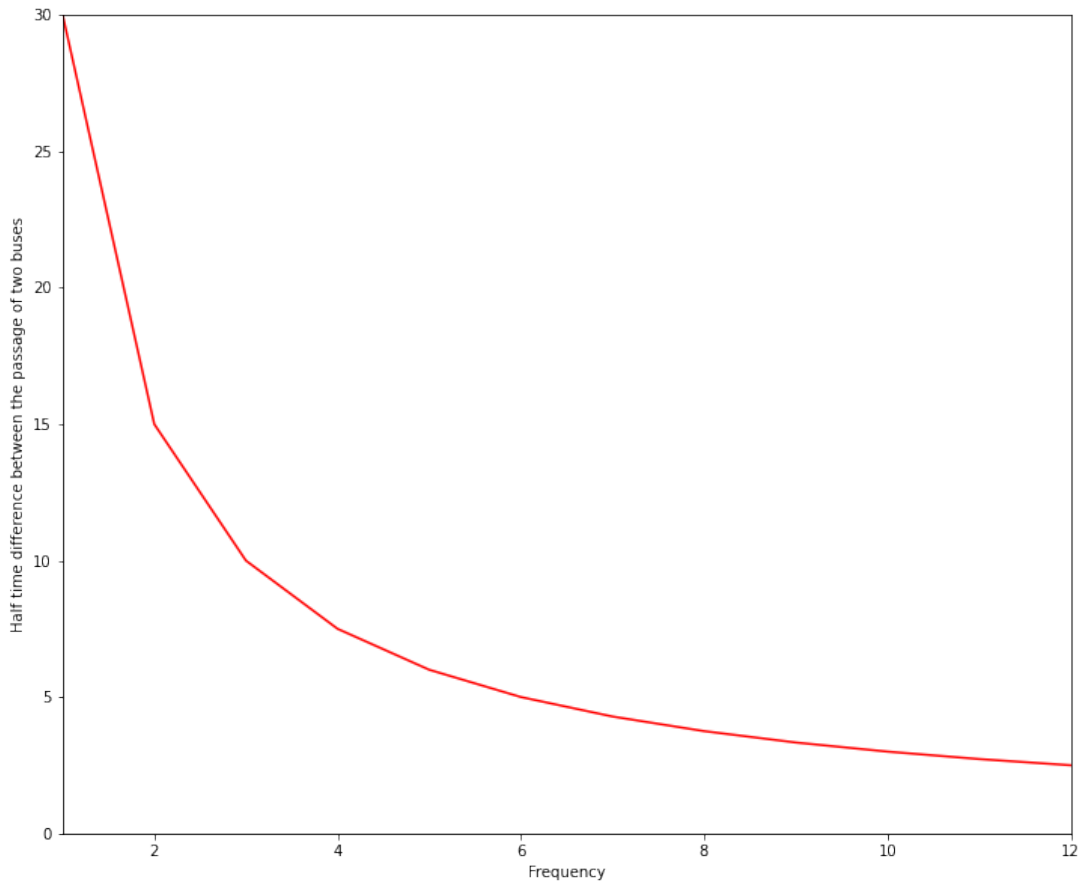


Figure 4.3 – Piecewise linear function to model the half time difference between the passage of two buses on a line as a function of its assigned frequency

function if we use boolean variables. Nevertheless, this function has flaws with respect to passenger behaviors.

Newell [1971] introduce a difference for low frequency services, where passengers adjust their arrival time at the stop in order to reduce their waiting time and for coordinated transfer time nodes, where the waiting time is not determined by the line frequency. Following that, Spiess and Florian [1989] present a model to estimate the waiting time of passengers, integrating passenger behavior for low frequency services. More recently, Yu et al. [2010] reuses this distinction by introducing choices about whether or not passengers should arrive at their stops uniformly, depending on the frequency of their lines. The authors recommend to use a fixed waiting time for low-frequency lines as they suppose that the passengers who use these lines adapt their arrival time at the bus stop to the line schedule. For high-frequency lines, they recommended to set waiting time to half the

headway interval. Based on Quinet [2013], Brethomé et al. [2018] reuse this distinction and set a limit between these two bus lines types at an operating frequency of 6 vehicles per hour.

Based on Newell [1971], Spiess and Florian [1989], Yu et al. [2010] and Brethomé et al. [2018], we assume an expected waiting time to be equal to half the time between two buses for high frequency lines (more than six buses per hour) while for low frequency lines (five buses per hour or less), the expected waiting time is set to five minutes. On Figure 4.4, with the frequency of the bus lines ranging from 1 to 12, we define the expected waiting times ranging from 2 minutes 30 seconds to 5 minutes.

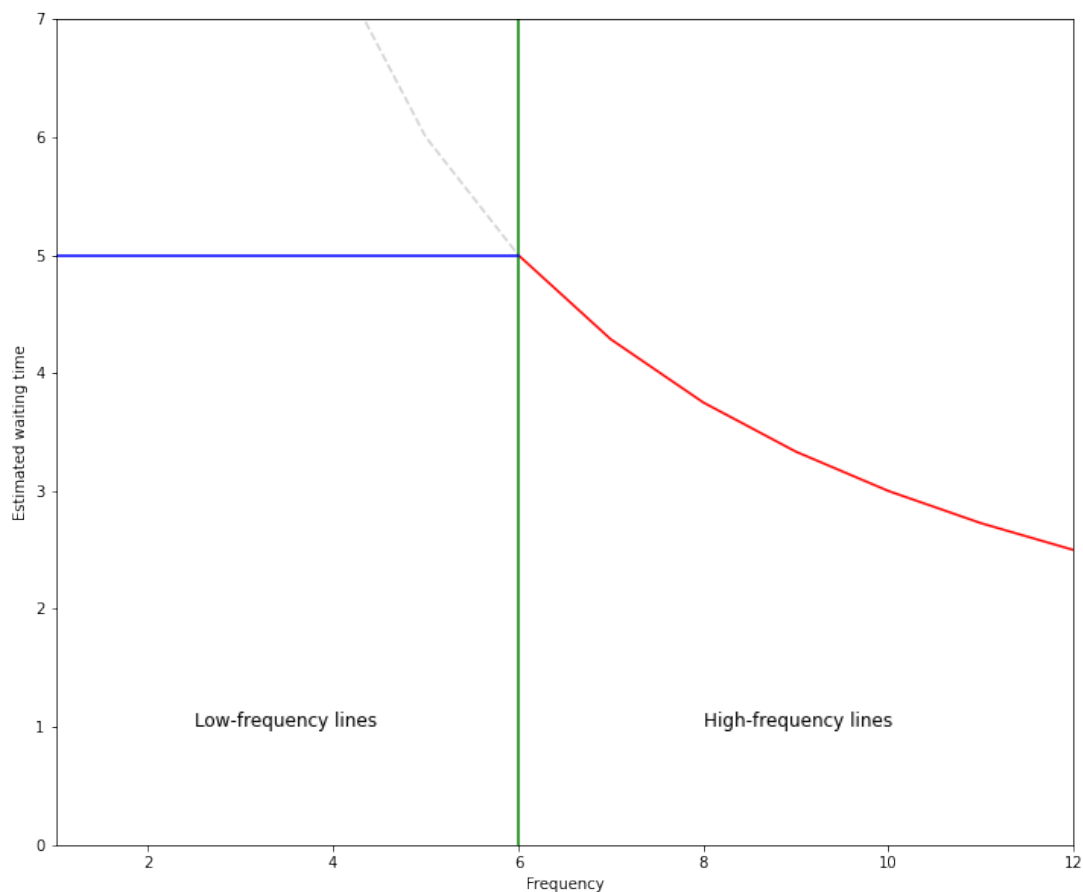


Figure 4.4 – Expected waiting times function considered in our work

Note that waiting times at transfers will be readjusted in the timetable development step after the frequency setting step (see e.g. Fouilhoux et al. [2016] on this topic). Hence, we assume that an expected waiting times can be calculated similarly for boarding at the first station of a path or for a transfer. We introduce a set Ψ containing all expected

waiting times. Ψ_f is associated with operating frequency Θ_f for all $f \in \mathcal{F}$. As Θ_f , Ψ_f is hence defined for $f \in \mathcal{F} = \{1, \dots, |\mathcal{F}|\}$. The waiting times Ψ_f of Figure 4.4 are defined in Table 4.1.

Table 4.1 – Expected waiting time values set according to line frequencies

Θ_f	0	1	2	3	4	5	6	7	8	9	10	11	12
Ψ_f	∞	5	5	5	5	5	5	4.29	3.75	3.33	3	2.73	2.5

To model service levels, we first presume that 100% of the demand has to be satisfied. Passenger satisfaction is then modeled according to two criteria: (1) path duration and (2) bus fill rate. The path duration criterion is based on the notion of *reference travel time* ξ_{st} . This reference travel time is defined for each OD pair and is estimated by the operator to provide passengers with a qualitative path duration from their origin to their destination. A first service level is used to enforce the duration of a passenger path not to exceed a given percentage α^{max} ($> 100\%$) of its OD reference travel time. Finally, to model passenger comfort, we define by ϕ_l the maximum bus filling rate of a bus line $l \in \mathcal{L}$. Table 4.2 summarizes these service level parameters.

Table 4.2 – Service level parameters

ξ_{st}	Reference travel time for (s, t) OD pair
α^{max}	Maximum deviation time of passengers from their reference travel time
ϕ_l	Maximum bus filling rate on bus line l during the considered time period

In the following section, we propose a MILP which integrates the following contributions:

- We set the sum of operating costs as the objective and not as a constraint.
- We use a path-based model to represent the path chosen by passengers on the network.
- We consider a heterogeneous bus fleet.
- We integrate constraints related to infrastructure restrictions.
- We use service level parameters for passengers.
- We integrate passenger assignment constraints assuming that passengers use their shortest paths, including in-vehicle travel and expected waiting times in the path duration model.

Table 4.3 summarizes the notation that is used to model the FSP.

Table 4.3 – Main notation for the FSP

$G = (V, A, \mathcal{L})$	Graph representing the infrastructure and topology of the bus network
V	Set of main stations
A	Set of road sections
\mathcal{L}	Set of bus lines
\mathcal{B}	Available bus types
\mathcal{F}_l	Set of possible frequencies for line $l \in \mathcal{L}$
\mathcal{P}	Set of passenger paths
\mathcal{P}_{st}	Set of passenger paths from s to t
\mathcal{L}_p	Set of bus lines taken by passenger path $p \in \mathcal{P}$
D	OD matrix for the considered time period
q_{st}	Quantity of passenger demand from s to t
Δ	Duration of the considered time period
o_b	Number of operable buses of type $b \in \mathcal{B}$
κ_b	Passenger capacity of buses of type $b \in \mathcal{B}$
s_a	Maximum number of buses on arc $a \in A$ for the considered period
r_l	One-way travel time of a bus on the line $l \in \mathcal{L}$
r_p	In-vehicle travel time of the passenger path $p \in \mathcal{P}$
$c_{l,b}$	Operating cost of the round-trip bus line $l \in \mathcal{L}$ with buses of type $b \in \mathcal{B}$
ξ_{st}	Reference travel time from s to t

1.2 The model

In this model, we use the following variables:

- $y_{l,f,b} = \begin{cases} 1, & \text{if line } l \in \mathcal{L} \text{ operates buses of type } b \in \mathcal{B} \text{ with a frequency of } \Theta_f \\ 0, & \text{otherwise} \end{cases}$
- $\gamma_p \in [0, 1]$, the percentage of the OD going from the origin to the destination of the passenger path $p \in \mathcal{P}$ that use this path.
- $x_p = \begin{cases} 1, & \text{if the passenger path } p \text{ is used} \\ 0, & \text{otherwise} \end{cases}$
- $z_{l,f} = \begin{cases} 1, & \text{if the combined frequency of the line } l \text{ with all bus types is equal to } \Theta_f \\ 0, & \text{otherwise} \end{cases}$
- $w_p \in \mathbb{R}_+$, equal to the product of x_p and $\sum_{l \in \mathcal{L}_p} \sum_{f \in \mathcal{F}} \Psi_f \times z_{l,f}$, meaning the waiting time when using the passenger path $p \in \mathcal{P}$.

FSP

$$\min \sum_{l \in \mathcal{L}} \sum_{f \in \mathcal{F}} \sum_{b \in \mathcal{B}} \Theta_f \times y_{l,f,b} \times c_{l,b} \quad (4.1)$$

$$\text{s.t.} \quad \sum_{f \in \mathcal{F}} y_{l,f,b} = 1 \quad \forall l \in \mathcal{L}, \forall b \in \mathcal{B} \quad (4.2)$$

$$\sum_{p \in \mathcal{P}_{st}} \gamma_p = 1 \quad \forall (s,t) \in D \quad (4.3)$$

$$\sum_{f \in \mathcal{F}} z_{l,f} = 1 \quad \forall l \in \mathcal{L} \quad (4.4)$$

$$\sum_{f \in \mathcal{F}} \sum_{b \in \mathcal{B}} y_{l,f,b} \times \Theta_f = \sum_{f \in \mathcal{F}} z_{l,f} \times \Theta_f \quad \forall l \in \mathcal{L} \quad (4.5)$$

$$\sum_{l \in \mathcal{L}} \sum_{f \in \mathcal{F}} 1.2 \times r_l \times y_{l,f,b} \times \Theta_f \leq o_b \times \Delta \quad \forall b \in \mathcal{B} \quad (4.6)$$

$$\sum_{l \in \mathcal{L} \mid a \in l} \sum_{b \in \mathcal{B}} \sum_{f \in \mathcal{F}} y_{l,f,b} \times \Theta_f \leq s_a \quad \forall a \in A \quad (4.7)$$

$$\sum_{f \in \mathcal{F}} \sum_{b \in \mathcal{B}} y_{l,f,b} \times \Theta_f \times \phi_l \times \kappa_b \geq \sum_{(s,t) \in D} \sum_{\substack{p \in \mathcal{P}_{st}: \\ (a,l) \in p}} \gamma_p \times q_{st} \quad \forall l \in \mathcal{L}, \forall a \in l \quad (4.8)$$

$$\gamma_p \leq x_p \quad \forall p \in \mathcal{P} \quad (4.9)$$

$$r_p \times x_p + w_p \leq \xi_{st} \times \alpha^{max} \quad \forall (s,t) \in D, \forall p \in \mathcal{P}_{st} \quad (4.10)$$

$$\sum_{l \in \mathcal{L}_p} \sum_{f \in \mathcal{F}} \Psi_f \times z_{l,f} \leq w_p + (1 - x_p) \times \Psi_1 \times |\mathcal{L}_p| \quad \forall p \in \mathcal{P} \quad (4.11)$$

$$x_p \times r_p + w_p \leq r_{p'} + \sum_{l \in \mathcal{L}_{p'}} \sum_{f \in \mathcal{F}} \Psi_f \times z_{l,f} \quad \forall (s,t) \in D, \forall p, p' \in \mathcal{P}_{st} \quad (4.12)$$

$$y_{l,f,b} \in \{0, 1\} \quad \forall l \in \mathcal{L}, \forall f \in \mathcal{F}, \forall b \in \mathcal{B}$$

$$z_{l,f} \in \{0, 1\} \quad \forall l \in \mathcal{L}, \forall f \in \mathcal{F}$$

$$\gamma_p \in [0, 1] \quad \forall p \in \mathcal{P}$$

$$x_p \in \{0, 1\} \quad \forall p \in \mathcal{P}$$

$$w_p \in \mathbb{R}_+ \quad \forall p \in \mathcal{P}$$

The objective function (4.1) minimizes the total estimated operating costs for the operator. Constraints (4.2) state that each line $l \in \mathcal{L}$ must be associated with one frequency Θ_f for each bus type $b \in \mathcal{B}$. Note that a frequency 0 must exist in Θ in order to allow a line not to be operated or not to be associated with a type of bus during a time period. Constraints (4.3) enforce the totality of each OD pair demand to be dispatched on

their respective passenger paths. *Constraints (4.4)* and (4.5) require that the combined frequency of a line for all of its operated bus types be defined in the \mathcal{F} set. *Constraints (4.6)* ensure that the number of buses used for operation does not exceeds the number of available buses of each type. To account for turnaround time, a factor of 20% is added to the riding time of a bus line. Turnaround time typically refers to the time required for a bus to complete a trip on one line and then begin a trip on the same line in the opposite direction, as well as any time required for bus repositioning between lines. *Constraints (4.7)* state that the number of buses of each type on a street section during the period is lower than the saturation limit. *Constraints (4.8)* make sure that a sufficient frequency is assigned to each bus line according to the passenger flow traveling on the network and the buses capacities. *Constraints (4.9)* ensure that the passenger path $p \in \mathcal{P}$ is used by a part of the flow only if this path is a shortest path in terms of path duration. *Constraints (4.10)* state that all passengers should have a path duration of at most α^{max} times their reference travel time. *Constraints (4.11)* calculate the waiting time on each path. In this constraint, $\Psi_1 \times |\mathcal{L}_p|$ refers to the longest possible waiting time (Ψ_1 minutes times the number of lines used by taking the passenger path). *Constraints (4.12)* enforce a passenger to take its fastest possible path with chosen frequencies. This constraint is a mono level formulation of the bi-level FSP which is discussed below.

Formulation of shortest path constraint (4.12)

Let us first introduce the bi-level formulation which enforces each path used by an OD pair to be a shortest path for this OD pair according to its path duration. For this purpose, we define the upper-level model as the model 1.2 without constraint (4.12). The lower-level problem is defined for each OD pair $(s, t) \in D$ and consists in selecting the passenger path minimizing its path duration.

$$\min \sum_{p \in \mathcal{P}_{st}} x_p \times (r_p + \sum_{l \in \mathcal{L}_p} \sum_{f \in \mathcal{F}} \Psi_f \times z_{l,f}) \quad (4.13)$$

$$\text{s.t.} \quad \sum_{p \in \mathcal{P}_{st}} x_p = 1 \quad (4.14)$$

$$x_p \in \{0, 1\} \quad \forall p \in \mathcal{P}$$

Based on Dempe and Mefo Kue [2017] and Schmidt and Beck [2021], we introduce the

optimal value function f defined as:

$$f_{st}(z) := \min\left\{ \sum_{p \in \mathcal{P}_{st}} x_p \times (r_p + \sum_{l \in \mathcal{L}_p} \sum_{f \in \mathcal{F}} \Psi_f \times z_{l,f}) : \sum_{p \in \mathcal{P}_{st}} x_p \geq 1, x_p \in \{0, 1\} \right\} \quad (4.15)$$

$$f_{st}(z) := \min_{p \in \mathcal{P}_{st}} \left\{ r_p + \sum_{l \in \mathcal{L}_p} \sum_{f \in \mathcal{F}} \Psi_f \times z_{l,f} \right\} \quad (4.16)$$

By using variable w_p defined by constraint (4.11), the path duration of the passenger path p can be expressed as $x_p \times r_p + w_p$. The constraint enforcing the passengers to use passenger paths with the smallest path duration possible can hence written as:

$$f_{st}(z) \geq x_p \times r_p + w_p \quad \forall (s, t) \in D, \forall p \in \mathcal{P}_{st} \quad (4.17)$$

This finally leads to the definition of constraint (4.12).

2 Solving process

The number of paths that can be taken by passengers can quickly become huge as the size of the graph and the number of bus lines operated increases. In addition, the cardinality of constraints (4.12) is quadratic in the number of paths that serve the same OD pair. It is therefore interesting to propose a path selection process in order to select paths that are interesting for the passengers and lead to reasonable operational costs for the operator. This section therefore focuses on this process.

2.1 Generating a set of passenger paths

This subsection focuses on the generation of a set of passenger paths, named Ω , based on a filtering and a dominance step. For each passenger path we define two notions: the *minimum path duration* and the *maximum path duration*. The *minimum path duration* t_p^{min} of a passenger path p is defined as the sum of the in-vehicle travel time and the lowest possible waiting time at departure and at each transfer. The *maximum path duration* t_p^{max} of a passenger path p is defined as the sum of the in-vehicle travel time and the highest possible waiting time at departure and at each transfer. The lowest and highest possible waiting time values are defined according to Ψ_2 and $\Psi_{|F|}$ the extrema values of the expected waiting time Ψ for an operated bus line. These minimum and maximum

path durations can hence be written as:

$$t_p^{min} = r_p + |\mathcal{L}_p| \times \Psi_{|F|} \quad (4.18)$$

$$t_p^{max} = r_p + |\mathcal{L}_p| \times \Psi_2 \quad (4.19)$$

Filtering step

The filtering step removes paths which cannot respect the maximum deviation time of passengers from their reference travel time. For each passenger path $p \in \Omega$, connecting nodes s and t , we evaluate if the minimum path duration t_p^{min} on p deviates by less than α^{max} from its reference travel time ξ_{st} . Mathematically: $t_p^{min} \leq \alpha^{max} \times \xi_{st}$. If this condition is not verified, the passenger path p is removed from the set Ω .

Dominance step

The dominance step evaluates if the minimum path duration t_p^{min} of a path p is lower than the maximum path duration t_{st}^{max*} of all passenger paths connecting the same origin and destination nodes. Hence we have $t_{st}^{max*} = \min_{p \in \mathcal{P}_{st}} (t_p^{max}, \alpha_{max} \times \xi_{st})$. Let us give an example where an OD pair would have several passenger paths and analyze the evaluation of each path. In this experiment, suppose the minimum transfer time is 2 minutes and the maximum is 5 minutes. The passenger paths are defined in Table 4.4. In this table, we separate the paths into three columns according to their number of transfers. Paths that are direct are called *direct paths* and paths that are associated to one or two transfers are respectively called *one-transfer paths* and *two-transfers paths*. Moreover, each passenger path p is associated with an index, a in-vehicle travel time r_p , a minimum path duration t_p^{min} and a maximum path duration t_p^{max} . As the maximum path duration of the passenger

Table 4.4 – Example of passenger paths authorized and unauthorized for an OD pair

Direct paths				One-transfer paths				Two-transfers paths			
p index	r_p	t_p^{min}	t_p^{max}	p index	r_p	t_p^{min}	t_p^{max}	p index	r_p	t_p^{min}	t_p^{max}
p_{01}	10	12	16	p_{11}	5	9	15	p_{21}	2	8	17
p_{02}	12	14	17	p_{12}	9	13	19	p_{22}	6	12	21
p_{03}	14	16	19	p_{13}	13	17	23	p_{23}	10	16	25

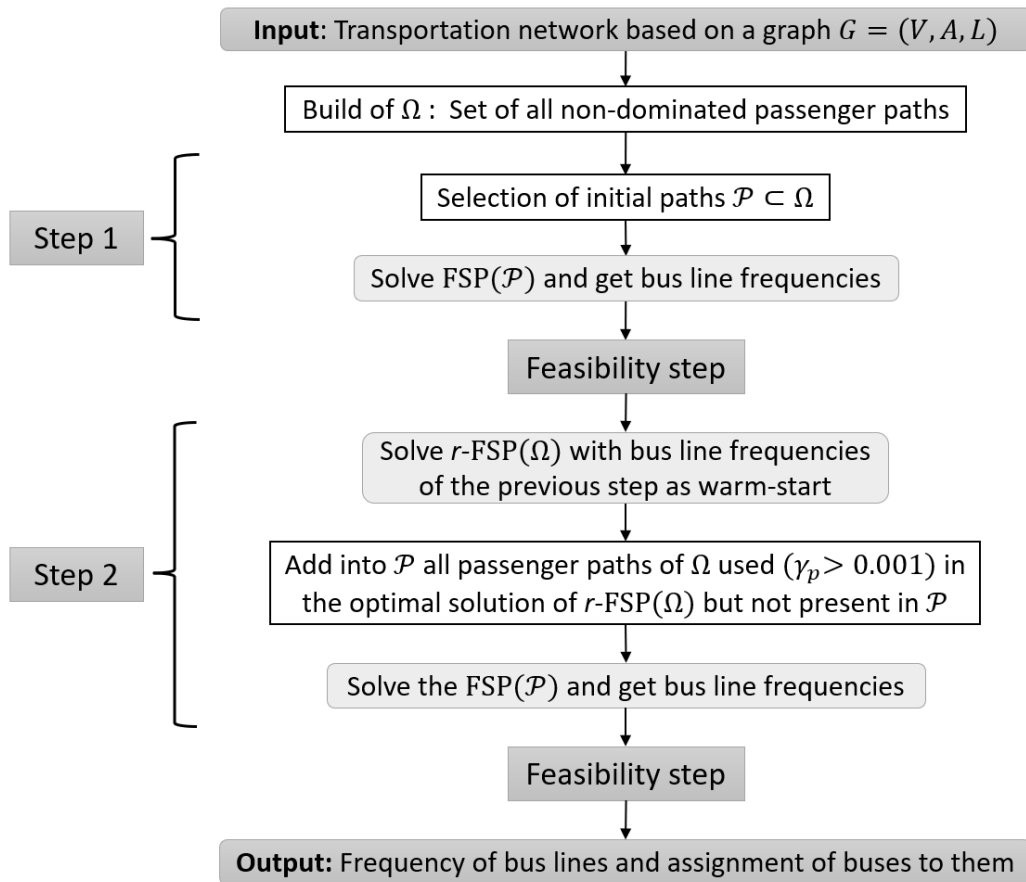
path p_{11} is smaller than the minimum path duration of passenger paths p_{03} , p_{13} and p_{23} . The passenger path p_{11} dominates the passenger paths with indexes p_{03} , p_{13} and p_{23} .

2.2 Passenger Path Selection Process

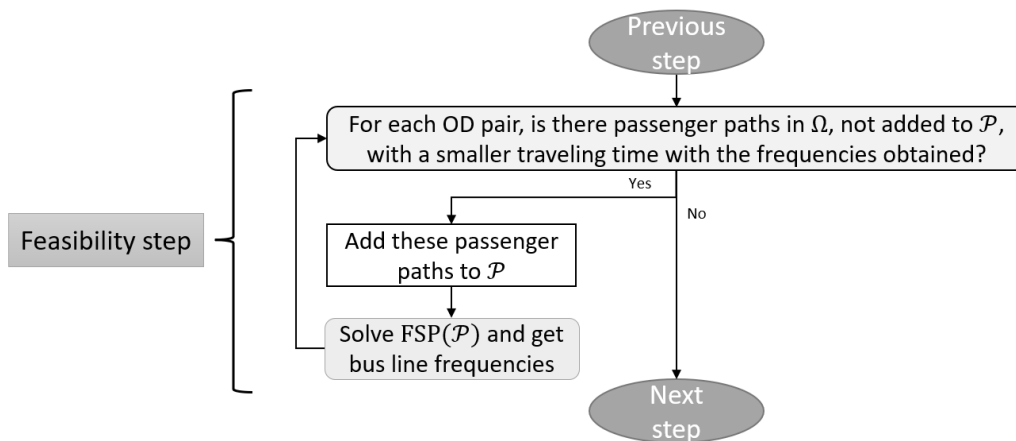
The FSP model in Section 1.2 minimizes operating costs by imposing operating constraints, service level based constraints and considering that the passengers take their shortest paths. Because of constraint (4.12) which has to be defined for each pair of paths that can serve the same OD pair, the solving time of the model highly depends on the number of paths. Hence, we propose the *Passenger Paths Selection Process* (PPSP) summarized in Figure 4.5. The PPSP consists of three steps:

- **Step 1: Initial frequency assignment.** The set Ω of non-dominated paths compatible with the targeted service levels is generated. See Subsection 2.1 for more details on this generation. Paths are then selected considering their minimum path duration. For each OD pair, if there exists some one-transfer paths in Ω that are shorter than the shortest direct paths already added to \mathcal{P} , the fastest of them is added to \mathcal{P} . Then, for each OD pair, if there are some two-transfer Ω paths that are shorter than the shortest direct and one-transfer paths already added to \mathcal{P} , the fastest of them is added to \mathcal{P} . Finally, $FSP(\mathcal{P})$ is solved to produce a solution denoted S .
- **Feasibility Step** Depending on the frequencies assigned to lines at the end of step 1, there may exist paths in $\Omega \setminus \mathcal{P}$ with shorter duration than those in \mathcal{P} for some OD pairs. If such a solution S use sub-optimal passenger paths it could therefore be considered unfeasible with respect to passenger choices. To restore the feasibility of the solution, we repeat a two-steps iterative process as long as the resulting solution is not feasible:
 1. For each (s, t) OD pair, for the frequencies of S , add to the set \mathcal{P} all paths from s to t in $\Omega \setminus \mathcal{P}$ of shorter duration than the paths used in S .
 2. Solve $FSP(\mathcal{P})$ to produce a new solution denoted S .
- **Step 2: Improve the feasible solution.** Solve $r - FSP(\Omega)$, the linear relaxation of $FSP(\Omega)$. The paths that are part of the solution and are used by a minimum percentage of the associated OD demand are added to \mathcal{P} . $FSP(\mathcal{P})$ is then solved.

Finally, a **feasibility step** is performed to make the final solution feasible. An example is given in Appendix B to illustrate the execution of this feasibility step on a small network.



(a) Overview of the Passenger Path Selection Process (PPSP)



(b) Feasibility step of the Passenger Path Selection Process (PPSP)

Figure 4.5 – Passenger Paths Selection Process

3 Experimentation using instances from the literature

In this section, we evaluate PPSP. The models are hence implemented with Julia and solved by Cplex 20.1 through the JuMP interface on a DELL R440 1U server with a 2.1 GHz Intel Xeon 6230 CPU and 192GB of RAM.

We are not aware of publicly shared instances for the FSP. As a result, we generated two bus networks based on two instances from the literature. See Appendix (C) for more details on these two instances. Data and results are shared in this thesis to allow further comparisons.

3.1 Experimentation settings

In order to be able to evaluate the PPSP to solve the FSP, we list different configurations. These configurations are described as follows:

- Use Step 1 followed by a feasibility step.
- Use PPSP with different acceptance threshold of a passenger path p in step 2:
 $\gamma_p = 1, \gamma_p > 0.1, \gamma_p > 0.01, \gamma_p > 0.001, \gamma_p > 0.0001$.
- Solve $FSP(\Omega)$.

	Time period duration	# bus fleet	Bus capacity	Frequency range	α^{max}	ϕ_l
<i>Dutch</i>	120 mns	1	600	[1, 24]	100%	100%
<i>Sioux – Falls</i>	120 mns	1	600	[1, 24]	100%	100%
<i>Poitiers_{SPPM}</i>	120 mns	1	100	[1, 24]	100%	21.97%

Table 4.5 – General parameters chosen to solve the FSP

We test these configurations on the Dutch and Sioux-Fall instances with the characteristics previously summarized in Chapter 3 in Tables 3.5 and 3.6. These characteristics are reminded and complemented in Table 4.5. For these two instances, the reference travel time ξ_{st} of each OD pair is arbitrarily chosen as the smallest path duration on the network, supposing an expected waiting time of 5 minutes at the first stop and for transfers.

We also test these configurations on the currently operated network of Poitiers described in Chapter 5 on the time period $6h - 8h$ and with the characteristics described

in Table 4.5. For this instance, the reference travel time ξ_{st} of each OD pair is defined according to the line frequencies offered by the local operator.

We define a network for each instance by solving the LPP-SL as presented in Chapter 3. Details concerning the achievement of these networks can be found in rows *CGEP* in Table 3.7 for the Dutch instance and in Table 3.9 for the Sioux-Falls instance.

3.2 Application of PPSP

Specification on the set Ω generated before and after filtering and domination steps for each instance is given in Table 4.7.

		# paths considered	Objective value	Solving time
<i>Dutch</i>	<i>PPSP(Step1 + feasibility)</i>	518	59000	3s
	<i>PPSP($\gamma = 1$)</i>	542	59000	3s
	<i>PPSP($\gamma > 0.5$)</i>	584	59000	4s
	<i>PPSP($\gamma > 0.1$)</i>	585	59000	4s
	<i>PPSP($\gamma > 0.01$)</i>	587	59000	4s
	<i>PPSP($\gamma > 0.001$)</i>	591	59000	4s
	<i>FSP(Ω)</i>	832	59000 (0%)	4s
<i>Sioux – Falls</i>	<i>PPSP(Step1 + feasibility)</i>	941	14.0414	864s
	<i>PPSP($\gamma = 1$)</i>	941	14.0414	955s
	<i>PPSP($\gamma > 0.5$)</i>	1069	14.0414	1037s
	<i>PPSP($\gamma > 0.1$)</i>	1090	14.0414	1182s
	<i>PPSP($\gamma > 0.01$)</i>	1299	14.0414	1740s
	<i>PPSP($\gamma > 0.001$)</i>	1301	14.0414	1724s
	<i>FSP(Ω)</i>	2969	15.2096 (14.94%)	5h (tl)
<i>Poitiers_{PPM}</i>	<i>PPSP(Step1 + feasibility)</i>	7224	7429	869s
	<i>PPSP($\gamma = 1$)</i>	7228	7429	1234s
	<i>PPSP($\gamma > 0.5$)</i>	7724	7371	1840s
	<i>PPSP($\gamma > 0.1$)</i>	7879	7286	2808s
	<i>PPSP($\gamma > 0.01$)</i>	8541	7286	4144s
	<i>PPSP($\gamma > 0.001$)</i>	8565	7286	4717s

Table 4.6 – Comparison of PPSP configurations on the generated networks with CGEP for Dutch, Sioux-Falls and Poitiers instances

		# paths	# direct	# 1-tr.	# 2-tr.
<i>Dutch</i>	Before filt. & dom. steps	8982	302	1868	6812
	After filt. & dom. steps	888	178	414	296
<i>Sioux – Falls</i>	Before filt. & dom. steps	52536	592	6236	45708
	After filt. & dom. steps	3724	464	1932	1328
<i>Poitiers_{PPM}</i>	Before filt. & dom. steps	542843	1702	32268	508873
	After filt. & dom. steps	110990	1445	18067	91478

Table 4.7 – Characteristics of the Ω set, before and after filtering and domination steps, for the Dutch, Sioux-Falls and Poitiers instances.

Note that starting PPSP directly from step 2 does not provide a feasible solution.

For the Dutch instance, we observe that the solving times are very low and do not allow to judge a potential time saving in terms of solving time using PPSP. Moreover, we observe that the solution found after executing step 1 and the feasibility step is the same as the one obtained by solving the FSP with the Ω path set. Hence operating costs found with the PPSP when varying the condition on γ make no change and lead to the same results. Using PPSP directly from step 2 does not provide a feasible solution whatever the value of time limit set.

For the Sioux-Falls instance, we observe that the time saving by using PPSP is significant on this network. No change in operating costs found with PPSP can be observed by varying the condition on γ . Using PPSP, an identical solution is obtained between 864 seconds and 955 seconds, depending on the acceptance condition on γ . This solution is better than the one obtained by solving FSP with all the paths with a time limit of 5h.

For the Poitiers instance, we notice that the path selection parameter, based on the γ value, has an important impact on the solving time. Different solutions are obtained with a solving time between 869 seconds and 4717 seconds, depending on the acceptance condition on γ . Moreover, we observe a real interest in executing PPSP from start to finish, step 2 included, because the result obtained at the output of the feasibility step after having solved step 1 is not as good as the ones obtained by executing PPSP. The setting of the acceptance condition on γ_p in step 2 is also important, the results obtained in output being dependent on its value. We thus observe that the best result is obtained by setting an acceptance threshold on $\gamma_p > 0.1$. Finally, note that we have no solution for the resolution of $FSP(\Omega)$, the memory required for it exceeding 10GB.

3.3 Frequencies Details for Dutch network

Index	Bus line	LPP-SL frequency	FSP frequency
Line n°1	[1, 21, 10, 9, 18, 17]	11	8/10
Line n°2	[2, 21, 4, 13]	7	6/8
Line n°3	[4, 19, 9, 18, 6, 21, 7, 20, 15]	7	5/6
Line n°4	[7, 21, 3, 19, 10]	12	11/12
Line n°5	[8, 5, 22, 21, 18, 9, 19, 3, 2, 11, 16]	7	8/9
Line n°6	[14, 12, 22, 13, 3, 19, 4, 21, 1, 23]	6	5/7
		Z = 60800	Z = 59000

Table 4.8 – Description of the lines frequencies obtained with PPSP on the network generated with CGEP for Dutch instance

For this network, we observe that the frequencies obtained by solving the FSP are close to those estimated by solving the LPP-SL. Lines n°1 and n°5 are slightly more used, while line n°3 is slightly less used. The estimated operating costs obtained when solving the FSP with PPSP (59000) are logically very close (-2.96%) to those estimated when solving the LPP-SL with CGEP (60800).

3.4 Frequencies details for Sioux-Falls network

Index	Bus line	LPP-SL frequency	FSP frequency
Line n°1	[2, 6, 8, 16, 17, 19, 15, 22, 23, 24, 13]	15	14 \ 11
Line n°2	[2, 1, 3, 12, 11, 10, 16]	15	20 \ 12
Line n°3	[4, 5, 9, 10, 11, 14, 23, 24, 13, 12]	14	7 \ 6
Line n°4	[4, 5, 6, 8, 7, 18, 16, 10, 15, 22, 21]	21	11 \ 10
Line n°5	[7, 18, 20, 19, 15, 14, 11, 4, 3]	12	7 \ 8
Line n°6	[12, 13, 24, 21, 22, 20, 19, 17, 16, 10, 9]	14	11 \ 16
Line n°7	[22, 23, 14, 11, 10, 15, 19, 17, 16]	8	6 \ 6
		z = 18.9924	z = 14.0414

Table 4.9 – Description of PPSP on the network generated with CGEP for Sioux-Fall instance

For this network, we observe that the frequencies obtained by solving the FSP with PPSP are more distant than those estimated by solving the LPP-SL with CGEP. Indeed, lines n°3, n°4, n°5 and n°7 are less exploited than estimated in the LPP-SL. We observe that the estimated operating costs obtained when solving the FSP are smaller by 26.07% than those estimated when solving the LPP-SL. The possibility for the lines to be non-symmetrical in terms of frequencies makes it possible to achieve these lower operating costs.

4 Conclusion

We proposed a model and a general methodology to solve the frequency setting problem. In this formulation, we minimize the operator cost while respecting operational constraints, service levels constraints for passengers and shortest paths constraints, modeling that each passenger can choose its shortest path on the network. The associated results can be used to refine frequencies on a network or to produce better cost and fleet estimations in a bus network design perspective. The problem being difficult to solve as the number of passenger paths increase, we proposed a methodology called Passenger Paths Selection Process (PPSP). This process progressively adds passenger paths to the model. It is based on three steps: (1) construction of an initial solution, (2) transformation of this solution into a feasible one, (3) improvement of this feasible solution. Computational results for two networks generated from instances of the literature were obtained and demonstrate the efficiency of our methodology.

CASE STUDY

Contents

1	Construction of the case study	98
1.1	Creating the network model	99
1.2	Generating OD Matrices	102
1.3	Categorizing the selected major stations	104
2	Filtering unrealistic lines	106
2.1	Topology-based filtering rule	106
2.2	Length and sinuosity based filtering rules	109
3	Restructuring the Vitalis Network	112
3.1	Bus lines	112
3.2	Service level evaluation	113
4	Designing bus networks with CGEP	115
4.1	Experimentation framework	115
4.2	Estimation of the maximum bus load parameter	116
4.3	Defining the adjustment coefficient	117
4.4	Five propositions of network reoptimization	118
4.5	Synthesis	122
5	Setting frequencies using PPSP	122
5.1	Experimentation framework	123
5.2	Operator data used in our experiments	123
5.3	Setting the bus filling rate parameter	124
5.4	Setting reference travel time deviation for passengers	125
5.5	Setting up the experimental framework	126
5.6	Setting the frequencies	127
5.7	Synthesis	131

To evaluate the viability of our approaches on a realistic case, we prospected users of the Heurès software. The Régie des Transports Poitevin (RTP) provided us with demand data collected from a large survey conducted on their network in 2014. This field survey, conducted on board the buses, provided information on various aspects of the transit system. It uses the IRIS territorial breakdowns. IRIS stands for “Ilots Regroupés pour l’Information Statistique” and is a system for dividing the territory of France into small geographic units to facilitate statistical analysis and data collection. The survey resulted in the definition of multiple time-periodic zone to zone OD matrices that contain the amount of demand between territorial breakdowns throughout the operating day. These demand matrices were used in this chapter to define a case study based on the bus network of Poitiers. Since our objective is to redesign the bus network without creating new stops or new links between the stops, the current Poitiers network is also taken as an input.

The operating costs per kilometer of the operator are not specified in this thesis. As specified in the Cerema report Cerema [2019] we use a cost per kilometer value of 4.36 €. To address the frequency setting problem, we assume this default value for standard buses in the heterogeneous bus fleet and adjust it by applying a cost increase of 15% for articulated buses.

We would like to acknowledge the “Régie des Transports Poitevin” (RTP) for sharing this data with us in the frame of this thesis and supporting the design of a more relevant decision support tool.

1 Construction of the case study

This subsection details the methods used to transform the surveyed data into an instance that can be used in a case study to test our algorithms. The first step consists in creating a graph of reasonable size that can be used as an input for our network design models. Following a common practice in us network design, we reduce the size of the problem by merging nodes in a judicious way. For this purpose, we select a set of stations that we will call *major stations* and that are of interest in terms of interchange points or the importance of passengers boarding and alighting.

1.1 Creating the network model

To model the network infrastructure, we first identify the existing stations on the Poitiers network. They will serve as potential nodes in our problem. Then, we identify the links between these nodes taking the street sections that are already used by buses.

To select the major stations, we consider three criteria: the geographic position of the station, its IRIS territorial breakdown, and the number of passengers boarding and alighting. The IRIS territorial breakdown provides a way to group nearby regions with similar socio-economic characteristics. Figure 5.1 shows the 36 zones in the Poitiers agglomeration defined by the IRIS territorial breakdown that are used in our analysis.

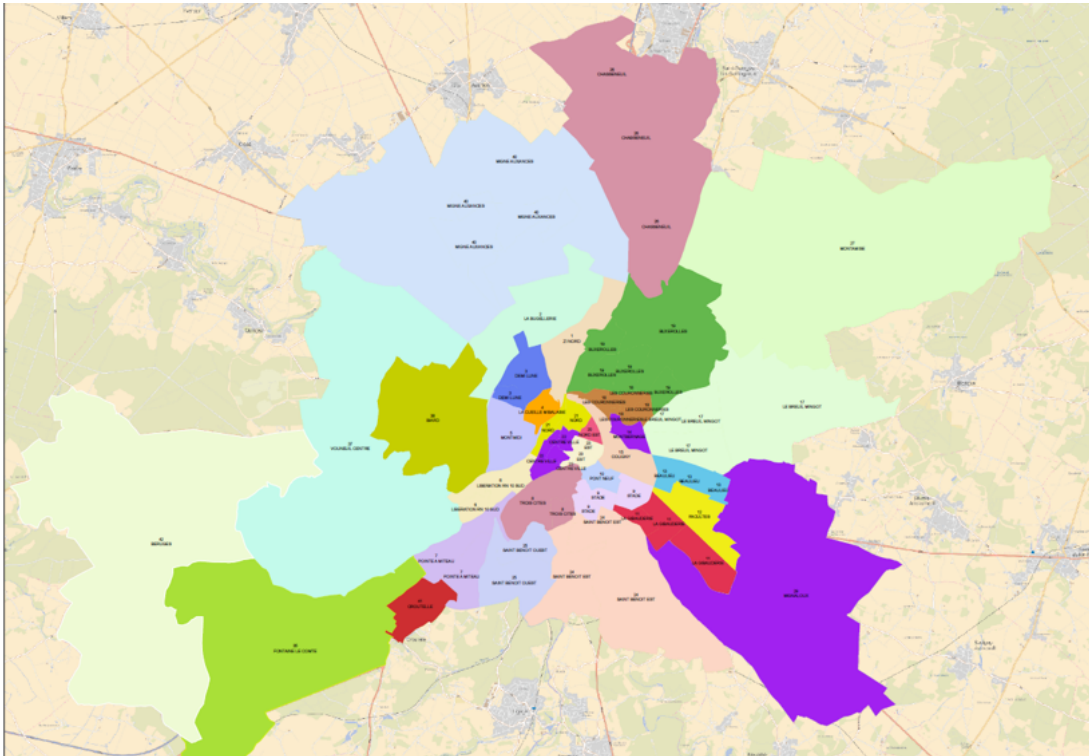
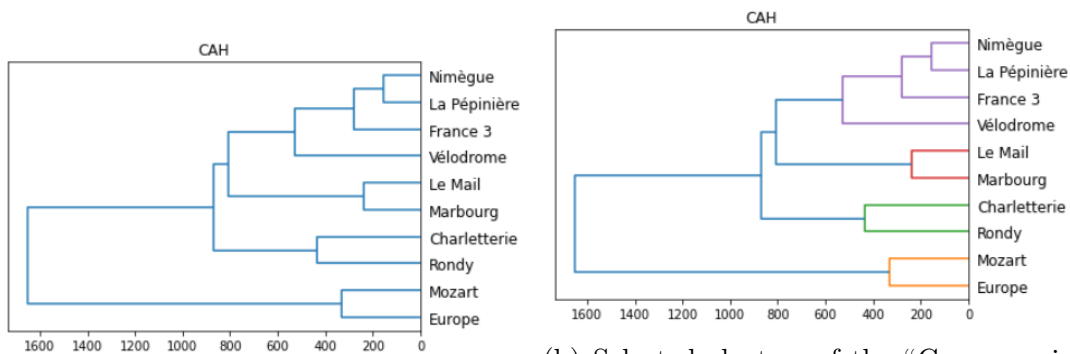


Figure 5.1 – The IRIS territorial breakdown of the agglomeration of Poitiers

To select the major stations, we proceed as follows: for each zone of the IRIS territorial breakdown, we apply a hierarchical clustering algorithm to group the stations based on their GPS coordinates. After creating the dendrogram, we can use it to partition the stations into a fixed number of clusters. This can be achieved using the "fcluster()" method, available in the Python's scipy library. We then select the most important station in each of the clusters established, based on the number of boarding/alighting

passengers at each station. These stations are the major stations in our network model.

Figure 5.2a presents a dendrogram obtained for the “Couronnerie” zone using a hierarchical clustering method. The height of the branches in the dendrogram represents the level of proximity between the stations. Figure 5.2b illustrates an example of clusters established when choosing to form four clusters. A cluster of stations is identified by its color.



(a) Hierarchical clustering result for the zone using the “fcluster()” method with stations of the “Couronneries” zone

(b) Selected clusters of the “Couronnerie” four clusters

By examining Figure 5.2b, we can see that when dividing the "Couronnerie" zone into four clusters, the clusters are defined as follows:

- Cluster n°1: {Nimègue, La Pépinière, France 3, Vélodrome}
- Cluster n°2: {Le Mail, Marbourg}
- Cluster n°3: {Charletterie, Rondy}
- Cluster n°4: {Mozart, Europe}

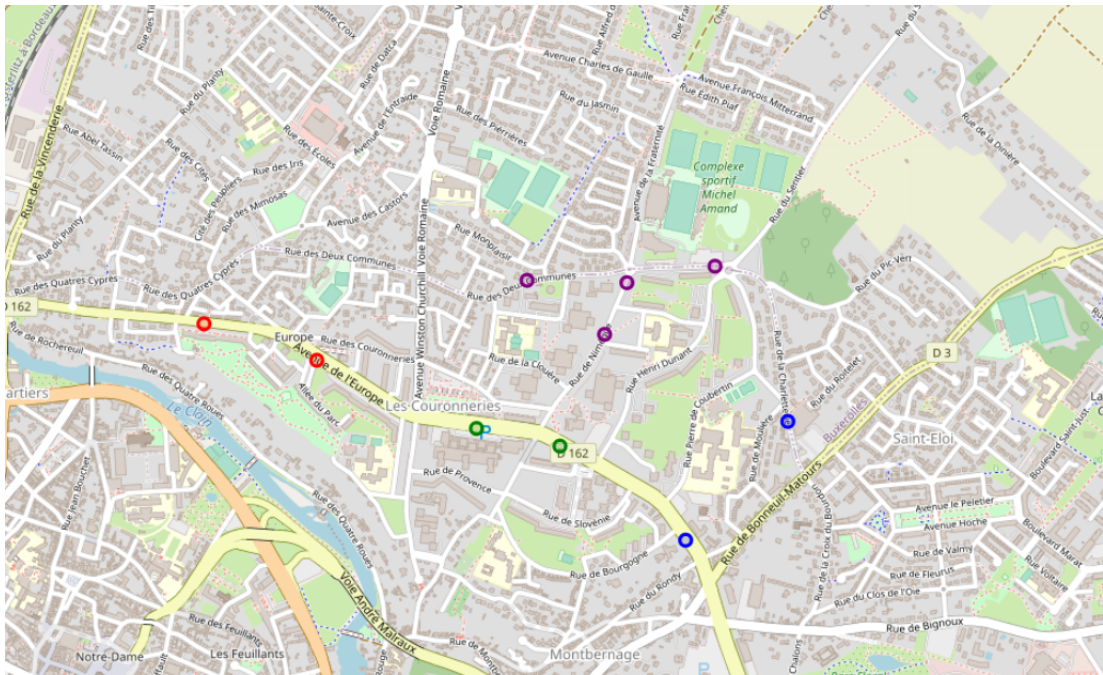


Figure 5.3 – Overview of the four clusters of the “Couronnerie” zone

Figure 5.3 gives an overview of the four clusters established for the “Couronnerie” zone. In our case, for each IRIS zone, we determine four clusters to identify four major stations per IRIS zone. However, an exception is made for small zones, where we determine three clusters to avoid creating clusters that are too small. We finally obtain a graph consisting of 80 nodes. To obtain information on the distance of street sections, we utilize OpenRouteService¹, a free and open-source routing service that uses OpenStreetMap data. We also use the Heurès software, supported by Lumiplan, to obtain in-vehicle travel times for the street sections. We define the arcs by including only the street sections currently used by at least one bus line in its route. The resulting graph consists of 218 arcs connecting the 80 major stations previously selected. This graph is depicted in Figure 5.4. On this graph, some of the nodes are represented in orange. These nodes are the potential terminus nodes, which are selected based on the currently operated terminus for each bus line.

1. <https://openrouteservice.org/>

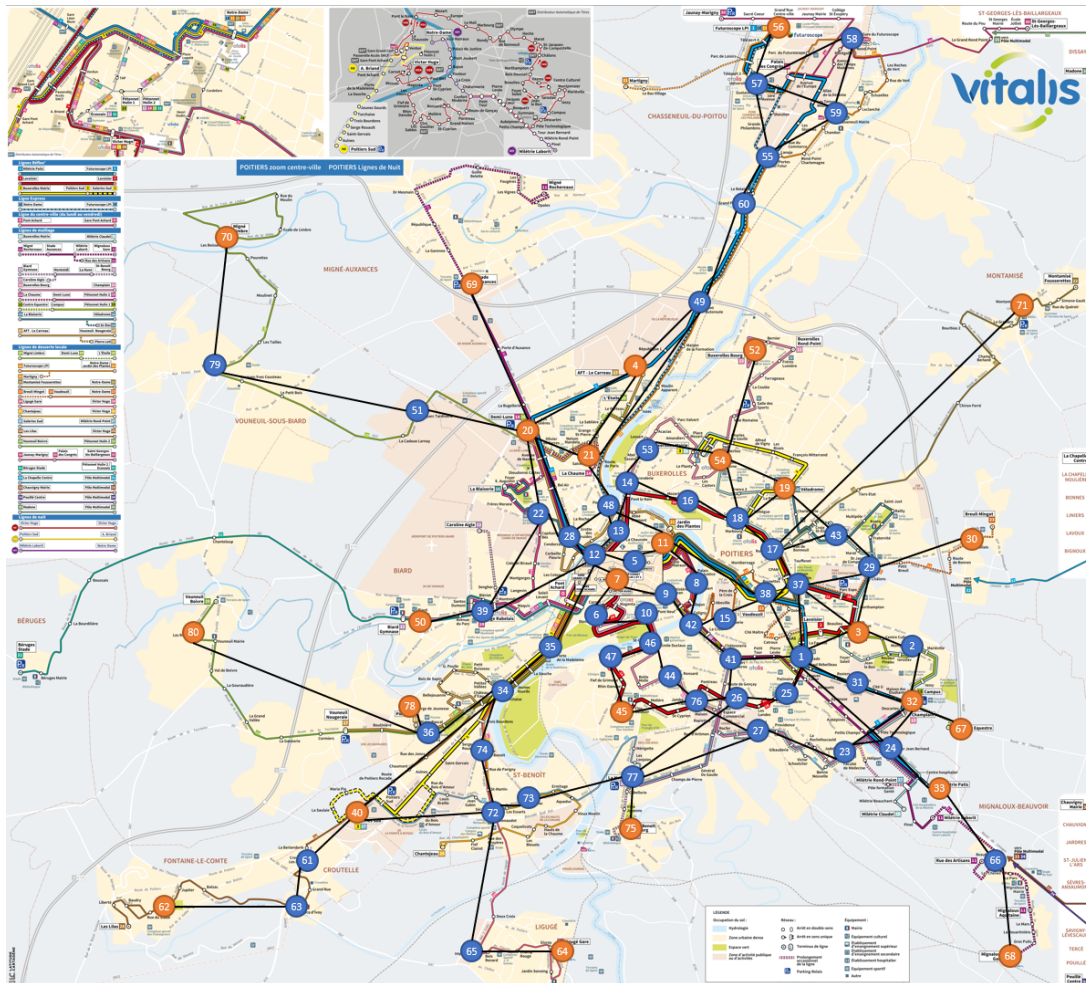


Figure 5.4 – Overview of the public transport graph generated for the case study.

1.2 Generating OD Matrices

To obtain “major station”-to-“major station” OD matrices from the zone-to-zone OD matrices provided by the RTP, we calculate a weight for each major station in each zone. To determine this weight, we use the boarding and alighting data for all stations.

The weight of a major station corresponds to the number of travelers who use the major station as their origin or destination in a given zone. We accomplish this by using the previously established clusters, where each cluster consists of a major station and its associated non-selected stations.

The demand from these non-selected stations is then assigned to the major station of its cluster, and the total number of alighting/boarding for the major station is calculated.

We obtain the weight of the major station by dividing the number of alighting/boarding for the major station by the total demand for the zone.

After computing the weight of each station significance percentages, we calculate the demand between pairs of major stations by multiplying the weights of the major stations by the estimated number of travelers between the two zones. These estimated numbers of travelers between two zones are specified in the zone to zone OD matrix provided by the operator, where each zone refer to an IRIS demarcation zone. . This leads to a “major station”-to-“major station” OD matrix that is useful for network optimization and restructuring.

Figure 5.2b illustrates the selected clusters of the “Couronnerie” zone using the “fcluster()” method with four clusters. The major stations are *Velodrome*, *Marburg*, *Rondy*, *Europe*. The assignment of non-selected stations to the major stations result in the following weights for each major station:

- Cluster n°1: stations *Nijmegen*, *La Pépinière*, *France 3* are assigned to station *Velodrome*. The weight of *Velodrome* within the zone is 7.89%.
- Cluster n°2: station *Le Mail* is assigned to station *Marbourg*. The weight of *Marbourg* within the zone is 55.13%.
- Cluster n°3: station *Charletterie* is assigned to *Rondy*. The weight of the major station *Rondy* within the zone is 26.15%.
- Cluster n°4: station *Mozart* is assigned to station *Europe*. The weight of the major station *Europe* within the zone is 10.83%.

Using the weights of major stations obtained for each zone, we can calculate the demand between pairs of major stations associated with those zones. The demand for a pair of major stations is computed by multiplying the weight of each major station and the amount of observed travel between the two zones.

Figure 5.5 summarizes the determination of demand for major station pairs in the “Couronneries” and “Demi-Lune” zones. The weight of a major station is denoted by $weight_x$, and $q^{zoneA,zoneB}$ represents the quantity of passengers demand between zones A and B.

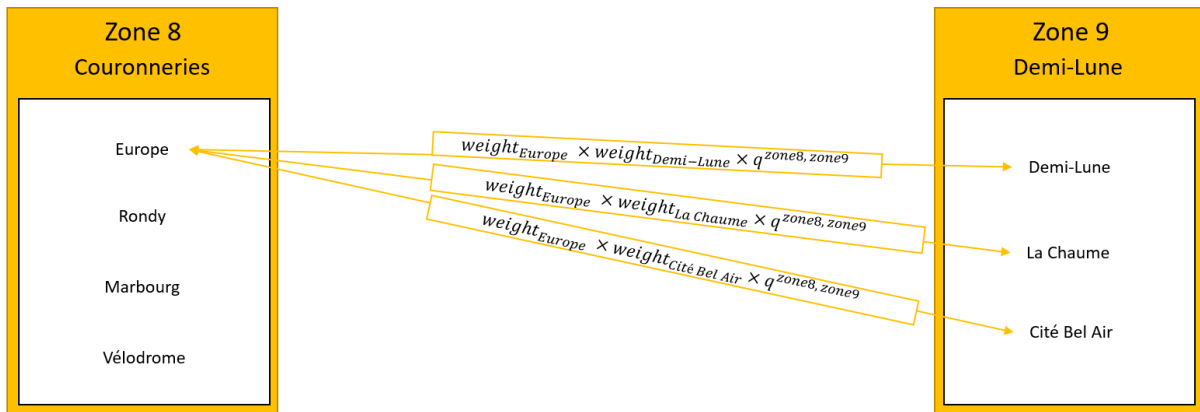
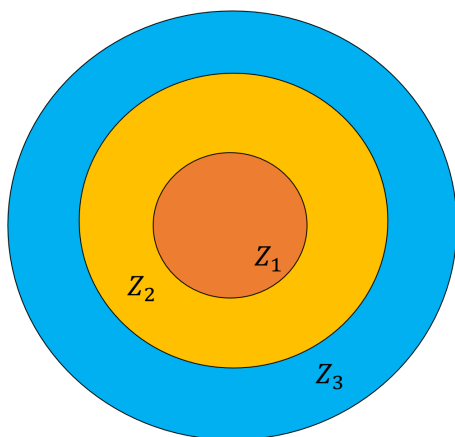


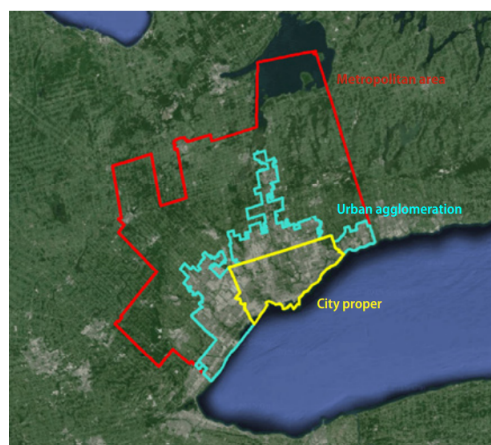
Figure 5.5 – Transforming zone-to-zone OD pairs to station-to-station OD pairs

1.3 Categorizing the selected major stations

We categorize stations in a network based on their location within the study area, using a categorization system based on three zones introduced by the United Nations report [WUP, 2018]. These three zones, namely, Z_1 , Z_2 , and Z_3 , correspond to the “city proper”, “urban agglomeration”, and “metropolitan area”, respectively, as illustrated in Figure 5.6a. The “city proper” (Z_1) corresponds to the administrative boundary of a city, while the “urban agglomeration” (Z_2) considers the extent of the contiguous built-up area, and the “metropolitan area” (Z_3) defines boundaries based on the level of economic and social interconnection of nearby areas. Figure 5.6b shows the urban definition boundaries for Toronto, Canada, according to the UN [WUP, 2018].



(a) Overview of the three zones used to categorize an urban territory

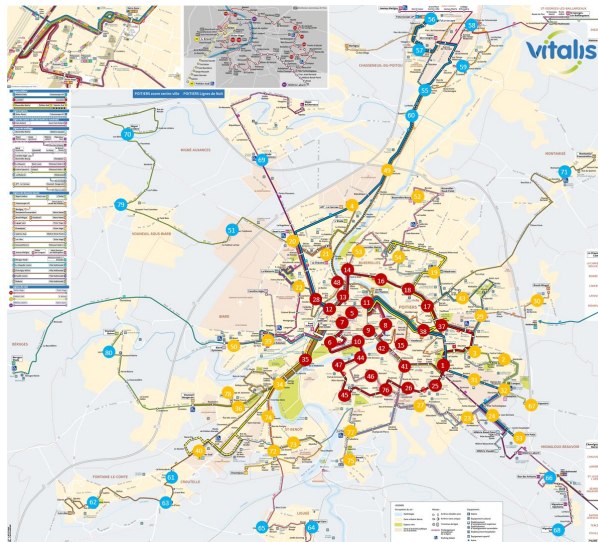


(b) Application of these three zones on the city of Toronto (Canada) according to the UN [WUP, 2018]

Additionally, according to Ritchie and Roser [2018], the term “urban agglomeration” denotes the population residing within a contiguous area with urban levels of density, regardless of administrative boundaries. This category typically includes the population of a city or town as well as that of the surrounding suburban areas that are contiguous and adjacent to the city’s boundaries.



(a) The Poitiers agglomeration divided into three zones



(b) Overview of the distribution of the major stations over the three zones of the Poitiers agglomeration

Figures 5.7a and 5.7b depict the Poitiers network, which comprises major stations classified into the three pre-defined zones.

2 Filtering unrealistic lines

In this section, we propose guidelines to filter unrealistic bus lines. These guidelines are based on the line typology of Faivre d’Arcier [2014] and on the examination of existing bus lines in the networks of Paris, Nantes, Rennes, Brest, Poitiers, and Lorient.

The purpose of these line typology guidelines is to enhance the topology of the proposed networks. The use of these typology rules serves two important purposes: first, it eliminates unrealistic bus lines from consideration; second, it helps reducing the line pool size and hence solving the LPP-SL during the execution of the CGEP.

2.1 Topology-based filtering rule

The first filtering rule in CGEP is designed to produce more realistic bus lines that are aligned with city policies. The rules used in this process are based on the classification system proposed by Faivre d’Arcier [2014] and are categorized based on the function and location of each line. The resulting classifications are major connection, meshing connection, major diffusion, and secondary diffusion. As a result, during the execution of CGEP, the bus line labeling algorithm filters out all bus lines that do not belong to any of these four classes. In our study of the city of Poitiers, we define the three zones that are represented on Figure 5.7a. The line classification system proposed in Faivre d’Arcier [2014] is adopted, where lines are classified into four categories: major connection, meshing connection, major diffusion, and secondary diffusion. The filtering rules are established as a set of zone sequences, which are determined by the bus stop locations along each line and their corresponding zone assignments. In the case study, the filtering rules are based on three zones for the agglomeration of Poitiers, as illustrated in Figure 5.7b. The integration of these typology rules occurs in CGEP at the end of the bus line generation algorithm as part of the solution to the bus line pricing problems. A bus line that does not meet the criteria for any of the four previously described classes is considered unrealistic and therefore filtered out of the generated line pool. This helps to ensure that the remaining lines are aligned with the established guidelines. Further details on this process can be found in the Subsection 2.2.

2.1.1 Major connection lines

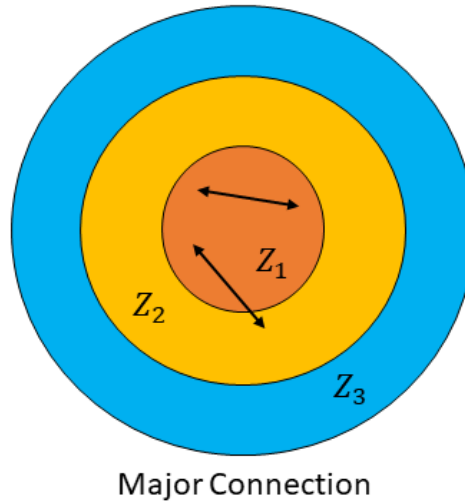


Figure 5.8 – Example of major connection lines

We define a major connection line as a bus line that primarily operates within the city proper (Z_1), and may also have a portion of its route within the urban agglomeration (Z_2). According to this definition, the Poitiers RTP bus network currently operates 10 major connection lines. Figure 5.8 shows examples of major connection lines on the three previously defined zones. For instance, a bus line with a sequence of bus stops (Z_1, Z_1, Z_2, Z_1) would be considered a feasible major connection line.

2.1.2 Meshing connection lines

We define a meshing connection line as a bus line that primarily operates within the urban agglomeration (Z_2), and may also have a portion of its route within the city proper (Z_1). In contrast to Faivre d’Arcier [2014], we allow a meshing connection line to have a short part of its route in the city proper in order to avoid being too restrictive when implementing these guidelines during the bus line generation process in CGEP. Currently, the Poitiers RTP bus network operates ten meshing connection bus lines, based on these criteria. Figure 5.9 shows examples of meshing connection lines within the three previously defined zones. As a demonstration, a bus line with a sequence of bus stops (Z_2, Z_2, Z_1, Z_2) would be considered a viable meshing connection line.

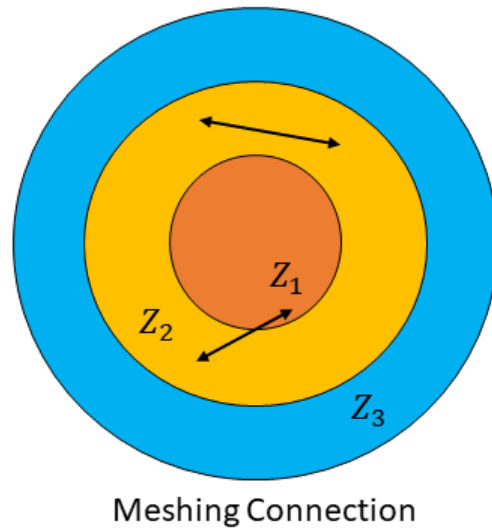


Figure 5.9 – Example of meshing connection lines

2.1.3 Major diffusion lines

We define a major diffusion line as a bus line that extends radially from the metropolitan area (Z_3) and urban agglomeration (Z_2) to the city proper (Z_1). Once reaching the city proper (Z_1), this line can either terminate there or continue on to either the urban agglomeration (Z_2) or metropolitan area (Z_3). Based on these criteria, the RTP bus network operating in the Poitiers agglomeration currently operates 9 major diffusion bus lines.

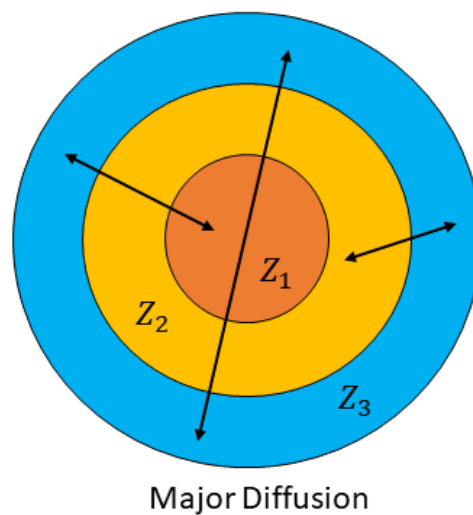


Figure 5.10 – Example of major diffusion lines

Figure 5.10 presents some examples of major diffusion lines. For instance, bus lines with sequences of bus stop locations such as $(Z_3, Z_3, Z_3, Z_2, Z_2, Z_1)$ and $(Z_3, Z_2, Z_1, Z_2, Z_2, Z_3)$ would be considered feasible major diffusion lines.

2.1.4 Secondary diffusion lines

We define a secondary diffusion line as a bus line connecting the different neighborhoods in the metropolitan area (Z_3). These lines primarily operate within the metropolitan area (Z_3) and may have a minor part of their route in the urban agglomeration (Z_2). Currently, the RTP-operated bus network in the Poitiers metropolitan area does not operate any secondary diffusion bus lines.

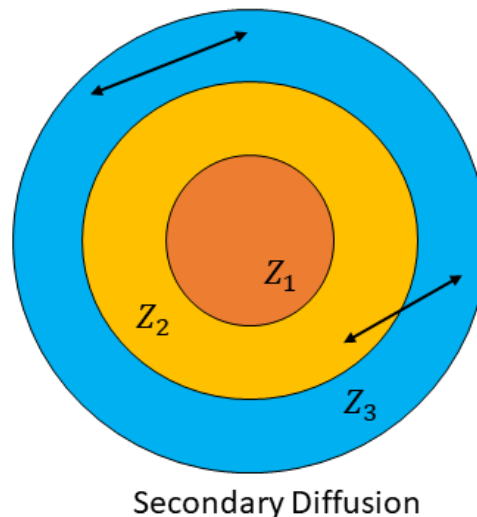


Figure 5.11 – Example of secondary diffusion lines

Figure 5.11 illustrates some examples of secondary diffusion lines within the three zones previously defined. For instance, a bus line that follows the sequence of bus stops $(Z_3, Z_3, Z_3, Z_2, Z_3)$ is considered a valid secondary diffusion line.

2.2 Length and sinuosity based filtering rules

The length of a bus line plays a crucial role in determining its feasibility and popularity among passengers. Lines that are too short may seem unappealing to users and may require transfers during their journey, while lines that are too long can lead to difficulty in maintaining schedules, causing missed runs and repositioning issues.

To address this, we establish two constraints on the length of bus lines. First, we specify a minimum and maximum length for each line. Specifically, we set the minimum length to 10 kilometers and the maximum length to 46 kilometers based on the shortest and longest lines in the existing Poitiers network, respectively. Second, we limit bus line sinuosity based on typology class for practicality. High sinuosity can negatively impact travel time and passenger appeal, so lower sinuosity is generally preferred for greater efficiency.

The **sinuosity index** is a measure of the deviation of a bus line from the shortest possible route between its termini. This is calculated as the ratio of the actual distance traveled by the line to the distance of the shortest path between the two termini. In our study, we define the sinuosity index as the ratio of the line distance to the distance of the shortest possible path, as opposed to the ratio of distance to bird’s eye view defined by Faivre d’Arcier [2014].



Figure 5.12 – Example of sinuosity index for the major diffusion bus line 79 of Nantes network

We illustrate this approach on Figure 5.12 using bus line 79 of the Nantes network that travels from the terminus “Orvault” to the terminus “Beauséjour”. According to the previously introduced typology classes, this line is classified as a major diffusion line. The length of this line is 9.5 kilometers, whereas the shortest possible distance is 7.4 kilometers. Hence, the sinuosity index of bus line 79 is equal to $\frac{9.5}{7.4} = 1.28$.

Faivre d’Arcier [2014] propose an alternative method of calculation based on the fact that passengers do not necessarily travel the entire length of a bus line. The author argue

that calculating from terminus to terminus may not always be relevant and introduce a new approach based on the *most significant station* served by the line. Typically, for main connection and meshing connection lines, this station is the one with the greatest number of connections. We define a **two-sections line** as a line running from $t1$ to $t2$ with its most significant station that is neither its terminus $t1$ nor its terminus $t2$. For two-sections lines, the sinuosity index can be calculated by averaging the two kilometer ratios (distance/shortest path distance) between this most important station and each terminus. This ensures a more general representation of the line's sinuosity.

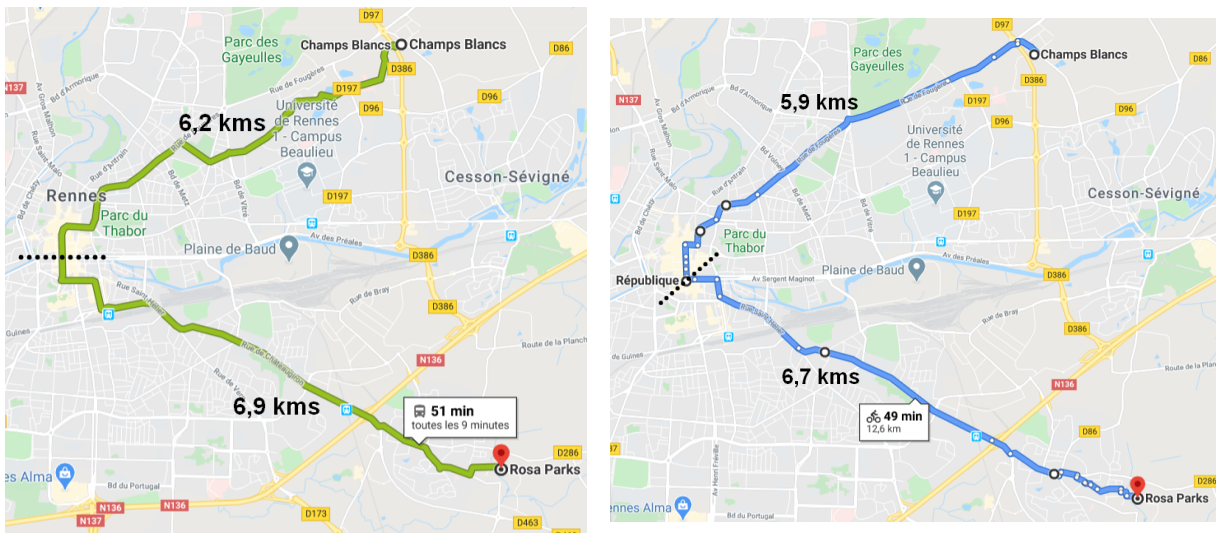


Figure 5.13 – Example of sinuosity index for the rapid transit bus line C1 of Rennes network

This approach is illustrated on the rapid transit bus line C1 in Rennes, as shown in Figure 5.13. This line connects the terminus *Champs Blancs* to the terminus *Rosa Parks*. The most significant station of this line is *République*. As a result, this line is considered a two-sections line. The length of the line between *Champs Blancs* and *République* is 6.2 kilometers, while the shortest possible distance is 5.9 kilometers. The length of the line between *République* and *Rosa Parks* is 6.9 kilometers, while the shortest possible distance is 6.7 kilometers. The sinuosity index of this line is calculated by taking the average of the two distances, $\frac{6.2}{5.9} = 1.05$ and $\frac{6.9}{6.7} = 1.03$, resulting in 1.04. Hence, the sinuosity index of the Rennes network's line C1 is equal to 1.04.

We propose to set the sinuosity ratios for our experiments based on the measurement of these indicators made on the existing network operated by RTP. The values are specified in Table 5.1.

Table 5.1 – Sinuosity ratios set for the experimentation on the Poitiers instance

	Connection		Diffusion	
	Major	Meshing	Major	Secondary
Sinuosity ratio	2	1.75	1.5	2

3 Restructuring the Vitalis Network

The aim of this section is to examine the impact of the restructuring of the Vitalis bus network in the Poitiers agglomeration, which took place in August 2015, on both the costs and the quality of service offered by RTP. The name “Vitalis” is used commercially by RTP to refer to the bus network it operates in the Poitiers area.

3.1 Bus lines

The pre-restructured Vitalis network consisted of 19 bus lines. According to the typology classification introduced by Faivre d’Arcier [2014], there were two major connection lines, seven meshing connection lines, and seven major diffusion lines. Additionally, three bus lines (lines 3, 7, and 8) did not belong to any of the four typology classes described in Subsection 2.1. Further details on the typology of the pre-restructured network can be found in Appendix D.1.

The post-restructured Vitalis network includes 28 lines. Conversely to the pre-restructuring network, all of the post-restructuring lines belong to one of the four typology classes described in Subsection 2.1. As a result of the restructuring, some areas that were previously not serviced by the pre-restructured network are now covered, including the territories of the municipalities of Béruges, Saint-Georges-les-Baillargeaux, La Chapelle-Moulière, Chauvigny, Puillé and Dissay. However, due to the absence of Origin-Destination (OD) data for these territories, we will only study 22 of the 28 bus lines of the post-restructured Vitalis network. These 22 bus lines cover the same territories as the 19 bus lines of the pre-restructured network.

Among the 22 considered lines of the post-restructuring network, we identify 3 major connection lines, 10 meshing connection lines, 9 major diffusion lines, and 0 secondary diffusion lines. Among the 22 considered bus lines, 16 existed prior to the restructuring, either in their current form or as a subsection of a former line. Additionally, according to

the line classification described in Subsection 2.1, we have identified 3 structuring lines, 1 major connection line, 10 meshing connection lines, 8 major diffusion lines, and no secondary diffusion lines.

	Pre-restructuring	Post-restructuring
Number of Lines	19	22
Total line length	490 kms	510 kms
Average line length	25.9 kms	23.1 kms
Median line length	27.1 kms	21.1 kms
Minimum line length	7 kms	10.5 kms
Maximum line length	49.5 kms	46 kms
Commercial kms operated	15,090 kms	13,600 kms
Cumulative km costs (base 4.36 euro/km)	65,792 €	59,296 €
CO2 emissions (base 1.2 kg/km)	18 108 kgs	16 320 kgs

Table 5.2 – Characterization of the pre- and post-restructuring networks

We will now examine the impact of the network restructuring on various metrics, including line length, number of lines, total kilometers operated, and cumulative kilometric costs.

To better understand the changes, we have compared the pre- and post-restructuring network characteristics for areas with available origin-destination (OD) demand data. These results are presented in Table 5.2.

We can see that there has been a decrease of approximately 10% in the cumulative kilometric costs between the pre-restructuring and post-restructuring Vitalis network for the areas with available OD demand data.

3.2 Service level evaluation

In this study, we assess the quality of service provided by the network using two service level criteria: (1) Passengers must have a route to travel from their origin to their destination without deviating more than a certain ratio from the shortest path. (2) A minimum percentage of passengers must be able to travel directly from their origin to their destination without any transfers. In the LPP-SL, we use two service level parameters, (α, β) . $\alpha \geq 1$ specifies the maximum deviation from the shortest in-vehicle travel time

that all passengers must not exceed. $\beta \in [0, 1]$ specifies the percentage of passengers that must have a direct path on the network.

Tables 5.3 and 5.4 and Figure 5.14 present the results of the service level parameters (α, β) for the Vitalis networks prior to and after restructuring, respectively, as determined by their respective network topology. These values are obtained by fixing the β parameter and then trying to decrease the α parameter as much as possible until the LPP-SL model is no longer feasible.

β	71%	72%	73%	74%	75%	76%	77%
α	1.60	1.68	1.80	1.86	2.10	2.25	2.72

Table 5.3 – Pair of service level parameters respected by the Vitalis pre-restructuring network

β	72%	73%	74%	75%	76%	77%
α	1.47	1.55	1.75	1.96	2.10	2.72

Table 5.4 – Pair of service level parameters respected by the Vitalis post-restructuring network

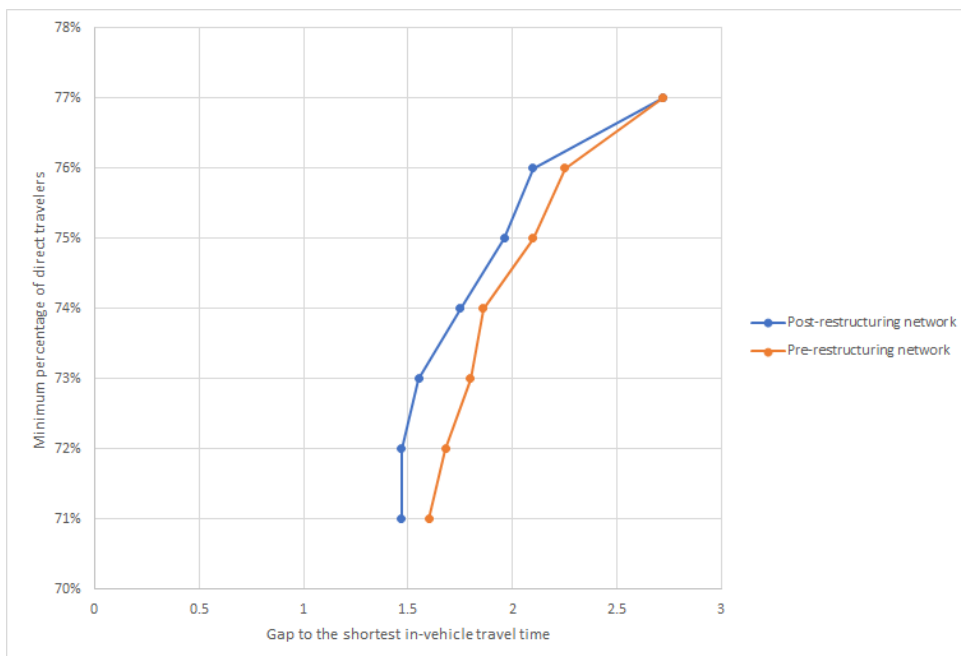


Figure 5.14 – Couples of service level parameters (α, β) introduced for LPP-SL and respected by Vitalis in pre- and post-restructuring networks

Figure 5.14 summarizes the results of the service level parameters. As can be seen on this figure, when considering the same percentage of passengers having a direct path on the network, the post-restructuring network has a lower maximum deviation from the shortest path travel time. As the street sections operated before and after restructuring are similar, these maximum deviation service level parameters can be compared directly. Therefore, based on the two established quality of service criteria, the post-restructuring network is of higher quality compared to the pre-restructuring network.

4 Designing bus networks with CGEP

This section highlights the application of CGEP to redesign the lines using the existing road infrastructure to ensure that the lines generated with CGEP are viable in terms of the available road infrastructure.

4.1 Experimentation framework

This experimentation consists of two steps. In the first step of the experiment, we integrate the topology of the post-restructuring network into the LPP-SL model. After that, we solve the LPP-SL with the aggregated OD matrix and obtain some frequencies. In the second step, we use the service level parameter pairs defined in Figure 5.5 to propose five different network re-optimization scenarios.

$\alpha \setminus \beta$	68%	74%	80%
2		x	
1.75	x	x	x
1.5		x	

Table 5.5 – Five different network re-optimization cases

In this experiment, two approaches are taken. (1) The maximum-spt-deviation (α) service level parameter is set to 1.75 and the minimum-rdc-percentage (β) service level parameter is given three values ranging from 68% to 80% (Table 5.5). (2) The minimum-rdc-percentage (β) service level parameter is set to 74% and three values of the maximum-spt-deviation α , ranging from 1.5 to 2, are tested (Table 5.5).

4.2 Estimation of the maximum bus load parameter

The setting of the maximum bus load is a topic that is not widely discussed in the literature. However, it has a significant impact on the performance of a real-world bus network. This subsection focuses on the role of the bus capacity parameter in LPP models and on its impact on the resulting network.

Through a review of three European bus manufacturers (Iveco, Mercedes-Benz, and Scania), it was found that standard city buses generally have a capacity of 90 to 100 passengers, with approximately 30 seats, while articulated urban buses have a capacity ranging from 150 to 160 passengers, with about 45 seats. However, buses are not always fully loaded throughout the day. Hence, we analyzed data from the partnership between Lumiplan and the RTP to determine the value that should be assigned to the bus capacity parameter.

In our LPP-SL model, as in most LPP models in the literature, the passenger capacity of buses is set with parameter κ , assuming that all buses operate at full capacity throughout the planning period. This can lead to some inaccuracies in estimating the actual number of passengers carried by the buses. Hence, κ should be viewed as a target for the average bus load during this period. This value serves as a rough indicator of congestion levels in buses, which is further refined in the frequency setting step.

Based on an analyze of the operator data, for a full day period from 6am to 9pm, the average number of passengers in a bus was found to be 18.9. We conduct a sensitivity analysis to examine the impact that this passenger capacity parameter κ can have on the optimized network. For this experiment, we consider lower and higher passenger capacity parameters of 14.2 and 23.6. A passenger capacity of buses of 14.2 corresponds to a 25% decrease in the passenger capacity from the estimated value of 18.9, while 23.6 corresponds to a 25% increase. The results in Table 5.6 reveal the significance of the maximum bus

κ	Estimated kms costs (base 4.36 €/km)	Estimated kms costs value evolution	CPLEX gap at timeout
14.2 (-25%)	55 751 €	+33.41%	8.20%
18.9	41 788 €	-	8.25%
23.6 (+25%)	33 471 €	-19.90%	7.77%

Table 5.6 – Influence of the maximum bus load parameter for Poitiers instance

load parameter on the optimized network. A 25% increase in the maximum bus load from 18.9 to 23.6 leads to a 19.88% decrease in the kilometric costs, with a similar value

of CPLEX gap at timeout. In contrast, a 25% decrease in the maximum bus load from 18.9 to 14.2 results in a 33.38% increase in the kilometric costs and an equal value of CPLEX gap at timeout. From these experiments, it seems that the objective value is highly sensitive to maximum bus load and that the latter must therefore be chosen with care.

K	CG Step			CE Step		
	# bus lines	# rdc paths	# indirect paths	# bus lines	# rdc paths	# indirect paths
14.2	550	3294	32238	577	7937	66236
18.9	618	3434	31650	639	7874	65210
23.6	515	3480	38776	527	7303	66742

Table 5.7 – Detail of the columns generated for Poitiers instance by varying the average bus load parameter

Table 5.7 provides some insight on the impact of this parameter on the method. We find that variations in the average bus load parameter have a reduced impact on the number of generated bus lines and passenger paths. Our experiments show that decreasing the average bus load from 18.9 to 14.2 resulted in a small decrease of 9.70% in the number of generated bus lines, from 639 to 577. However, there was almost no impact on the number of direct and indirect generated passenger paths. Similarly, when the average bus load was increased from 18.9 to 23.6, there was a small decrease of 17.53% in the number of generated bus lines (from 639 to 527) with no significant impact on the number of direct and indirect passenger paths.

4.3 Defining the adjustment coefficient

This subsection is focused on determining an adjustment coefficient for the LPP-SL in the context of the Poitiers instance. We input the current topology of Poitiers into our model and calculate estimated kilometric costs, which we refer to as the *reference network*. These estimated costs per kilometer may be more or less accurate, depending on the approximation of the infrastructure and passenger demand data used. To align these estimated costs with the operator’s actual costs, we calculate the ratio of the estimated costs for the reference network to the operator’s actual costs. We call this ratio the *adjustment coefficient*. This coefficient can then be applied to the output of CGEP to adjust the estimated kilometric costs and make them comparable to the operator’s actual

costs.

Based on our analysis using field survey data from RTP, the maximum bus load has been set to 18.9. You can find more information on this analysis in Subsection 4.2. The reference network, outlined in Table 5.8, is established by applying the LPP-SL model with service level parameters $(\alpha, \beta) = (1.75, 74\%)$ to an aggregated day (Monday through Saturday, from 6am to 9pm).

	Business day (6h - 21h)	
	Vitalis network	
	Estimated data	Operator data
Commercial kms operated	12 040 kms	13 600 kms
Estimated kilometric costs (base 4.36 €/km)	52 493 €	59 300 €
Estimated qty of CO_2 emitted (base 1.2 kg/km)	14 448 kg	16 320 kg

Table 5.8 – Comparison of the operator data with these estimated by the LPP model at iso topology

The difference in commercial kilometers operated is due to the approximations made in the calculation of arc kilometers using open route service. To adjust the estimated commercial kilometers and make them comparable to the commercial kilometers operated by the operator, we calculate the adjustment coefficient by taking the ratio of the total distances traveled in the operated bus network and the reference bus network. In the case of the Poitiers instance, the adjustment coefficient is approximately 1.13, calculated as $\frac{13,600}{12,040}$. We then multiply the estimated kilometric costs by this adjustment coefficient to obtain adjusted kilometric costs.

4.4 Five propositions of network reoptimization

Tables 5.9 and 5.11 present five reoptimization scenarios. The scenarios presented in Table 5.9 are based on a fixed minimum-rdc-percentage of 74% and varying the maximum-spt-deviation parameter, α , from 1.5 to 2. Conversely, the scenarios presented in Table 5.11 are based on a fixed maximum-spt-deviation parameter of 1.75 and varying the minimum-rdc-percentage, β , from 68% to 80%.

α	β		Business day (6h - 21h)	
			Proposed network	Proposed network adj.
2	74%	Estimated commercial kms	9 567 kms	10 811 kms
		Estimated kilometric costs (base 4.36 €/km)	41 711 €	47 133 €
		Estimated qty of CO_2 (base 1.2 kg/km)	11 480 kgs	12 972 kgs
1.75	74%	Estimated commercial kms	9 584 kms	10 830 kms
		Estimated kilometric costs (base 4.36 €/km)	41 788 €	47 219 €
		Estimated qty of CO_2 (base 1.2 kg/km)	11 500 kgs	12 995 kgs
1.50	74%	Estimated commercial kms	9 702 kms	10 963 kms
		Estimated kilometric costs (base 4.36 €/km)	42 302 €	47 801 €
		Estimated qty of CO_2 (base 1.2 kg/km)	11 642 kgs	13 155 kgs

Table 5.9 – Perspectives of network redesign with fixed minimum-rdc-percentage parameter

4.4.1 Impact of variations of the maximum-spt-deviation (α) service level parameter for a fixed minimum-rdc-percentage (β) service level

Table 5.9 presents three different scenarios of network design by varying the maximum-spt-deviation service-level parameter (α). The first column shows the estimated commercial kilometers, kilometric costs, and quantity of CO_2 obtained at the output of LPP-SL for each scenario, while the second column shows these same values adjusted using the adjustment coefficient specified in Subsection 4.3.

From these results, it is clear that adjusting the maximum-spt-deviation (α) service level parameter has limited impact on the estimated kilometric costs obtained through CGEP. By tightening this parameter from its original value of 1.75 to 1.5, the estimated kilometric costs increased by 1.23%. On the other hand, loosening the maximum-spt-deviation (α) parameter to 2 resulted in a decrease of 0.18% in the estimated kilometric costs. Table 5.10 highlights the results obtain for variations of the maximum-spt-deviation

α	β	CG			CE		
		# bus lines	# rdc paths	# indirect paths	# bus lines	# rdc paths	# indirect paths
2	74%	547	3378	49642	559	8811	121570
1.75	74%	618	3434	31650	639	7874	65210
1.5	74%	577	3125	1884	932	7237	33459

Table 5.10 – Columns generated for the different variants of the service level parameter α

(α) service level parameter on the execution of CGEP. The first three columns show the number of bus lines, direct passenger paths, and indirect passenger paths generated during the column generation step (CG) and the next three columns show the same information for the column enumeration step (CE).

It can be observed that by making the maximum-spt-deviation (α) service level parameter more restrictive, i.e. decreasing it from 1.75 to 1.5, results in a decrease in the number of generated passenger paths. This is because the passenger path labeling algorithm filters out paths based on their duration. As a result, the number of direct passenger paths decreases from 7874 to 7237 (-8.09%) and the number of indirect passenger paths decreases from 65210 to 33459 (-48.69%). On the other hand, the number of bus lines increases from 639 to 932 (+45.85%).

Conversely, making the maximum-spt-deviation (α) service level parameter less restrictive, i.e. increasing it from 1.75 to 2, leads to an increase in the number of generated passenger paths. The number of direct passenger paths increases from 7874 to 8811 (+11.90%), the number of indirect passenger paths increases from 56210 to 121570 (+116.28%), and the number of bus lines decreases from 639 to 559 (-12.52%).

It can be noted that reducing coefficient α implies finding a feasible set of passenger paths and lines to initialize the CGEP. This was done by modifying the existing lines manually for $\alpha = 1.5$. We did not reduce this parameter further for this reason.

4.4.2 Impact of variations of the minimum-rdc-percentage (β) service level parameter for a fixed maximum-spt-deviation (α) service level

Table 5.11 presents three different scenarios of network design by varying the minimum-rdc-percentage service-level parameter (β). The first column shows the estimated commercial kilometers, kilometric costs, and quantity of CO_2 obtained at the output of LPP-SL for each scenario, while the second column shows these same values adjusted using the adjustment coefficient specified in Subsection 4.3.

From these results, we observe that the minimum-rdc-percentage (β) service level parameter has varying impacts on the estimated kilometric costs. When the minimum-rdc-percentage is reduced from 74% to 68%, there is no change in the estimated kilometric costs. However, when the minimum-rdc-percentage is increased from 74% to 80%, there is a noticeable increase of 8.14% in the estimated kilometric costs.

In Table 5.12, we observe that, as the minimum-rdc-percentage (β) service level parameter is decreased from 74% to 68%, the number of bus lines generated increases from

α	β		Business day (6h - 21h)	
			Proposed network	Proposed network adj.
1.75	68%	Estimated commercial kms	9 584 kms	10 830 kms
		Estimated kilometric costs (base 4.36 €/km)	41 788 €	47 219 €
		Estimated qty of CO_2 (base 1.2 kg/km)	11 500 kgs	12 995 kgs
1.75	74%	Estimated commercial kms	9 584 kms	10 830 kms
		Estimated kilometric costs (base 4.36 €/km)	41 788 €	47 219 €
		Estimated qty of CO_2 (base 1.2 kg/km)	11 500 kgs	12 995 kgs
1.75	80%	Estimated commercial kms	10 364 kms	11 711 kms
		Estimated kilometric costs (base 4.36 €/km)	45 186 €	51 060 €
		Estimated qty of CO_2 (base 1.2 kg/km)	12 437 kgs	14 054 kgs

Table 5.11 – Perspectives of network redesign with fixed maximum-spt-deviation parameter

α	β	CG			CE		
		# bus lines	# rdc paths	# indirect paths	# bus lines	# rdc paths	# indirect paths
1.75	68%	573	3031	38197	1095	9173	72788
1.75	74%	618	3434	31650	639	7874	65210
1.75	80%	466	3506	35168	491	7423	69940

Table 5.12 – Columns generated for the variation of the service level parameter β

639 to 1095, a growth of 71.36%. Additionally, the number of rdc-passenger paths generated also increases from 7874 to 9173, a rise of 16.50%. On the other hand, when the minimum-rdc-percentage (β) service level parameter is increased from 74% to 80%, the number of bus lines generated decreases from 639 to 491, a decrease of 23.16%. Additionally, the number of rdc-passenger paths generated decreases from 7874 to 7423, a decline of 5.73%.

It is also noted that there is no significant change in the number of indirect passenger paths generated, regardless of whether the minimum-rdc-percentage (β) service level parameter is decreased or increased.

4.5 Synthesis

We have proposed five bus networks by adjusting the service level parameters of maximum-spt-deviation (α) and minimum-rdc-percentage (β). Based on our results, we have determined that three out of the five proposed networks are cost-effective with respect to the selected service level parameters. These networks have a service level that is either equivalent to or superior to that of the current operator's network. In Table 5.13, we refer to these three networks as *A*, *B*, and *C*.

$\alpha \setminus \beta$	68%	74%	80%
2			
1.75		<i>A</i>	<i>C</i>
1.5		<i>B</i>	

Table 5.13 – Names of the three cost-effective networks based on service level parameters

In the following section, we will assign frequencies to the lines of these three identified networks.

5 Setting frequencies using PPSP

In this section, we will describe the use of the PPSP method to determine the bus line frequencies for the *A*, *B* and *C* networks defined in Section 4. This step allows us to refine the cost and service level estimations for each network. Indeed, as presented in Chapter 4 it takes account of several periods with different demand in the day and integrates a better choice model for the passengers as well as a better path duration estimation.

5.1 Experimentation framework

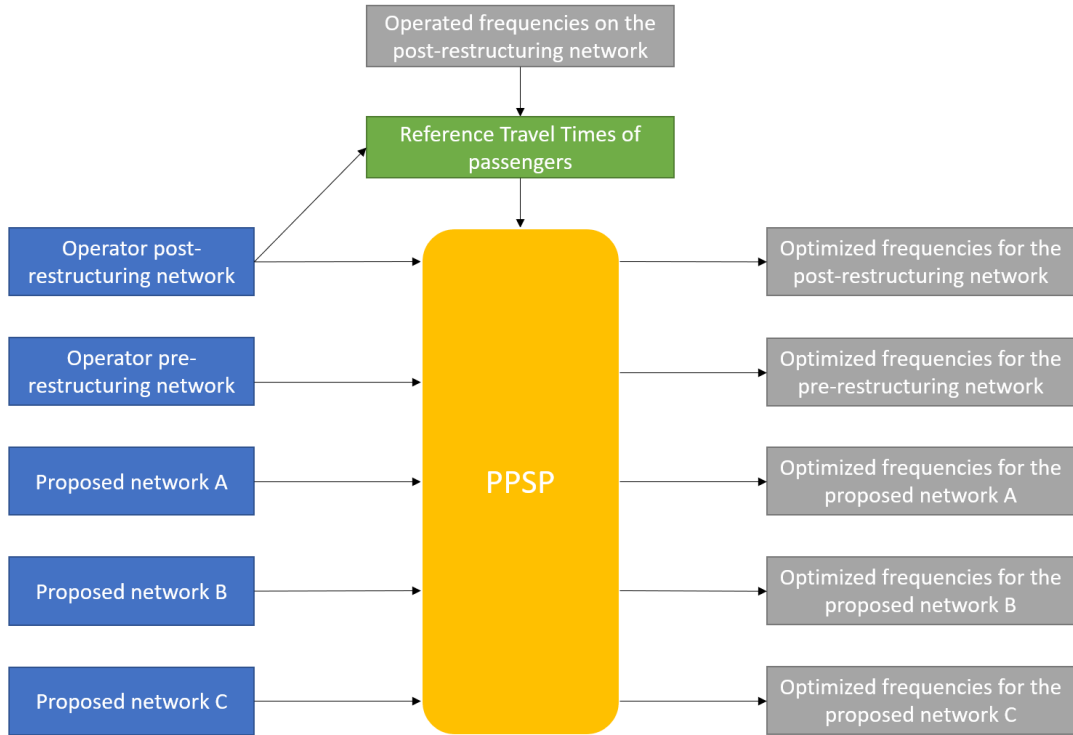


Figure 5.15 – Five line plans compared

We propose an experiment that aims at comparing the PPSP results obtained by solving CGEP for the pre-restructuring and post-restructuring networks of the operator, as well as for networks A , B , and C obtained by solving the CGEP. The process of the experiment is summarized in Figure 5.15. To estimate the passenger travel times in the PPSP, we use the travel times estimated in the post-restructuring network, based on the frequencies selected by the operator, as reference.

5.2 Operator data used in our experiments

For our experiments, we use the time periods defined by the RTP operator. The time periods, along with the associated number of OD pairs and passengers, are specified in Table 5.14. The passenger quantities are not necessarily integers, as they are calculated using the method described in Sub-section 1.2.

		# OD pairs	# passengers	Average # pass. per hour
Monday-Friday	6h-8h	1796	6499.12	3249.56
	8h-12h	2032	10535.27	2633.82
	12h-14h	1780	5970.09	2985.05
	14h-16h30	1956	8069.34	3227.74
	16h30-18h30	2103	8783.70	4391.85
	18h30-21h	1437	2936.67	1174.67

Table 5.14 – Specification of supposed passengers demand per considered time periods

Table 5.15 provides information on the RTP’s bus fleet. Information regarding the type and number of vehicles is obtained from the RTP operator, while data regarding the maximum bus load is observed from OMNIL (Observation de la Mobilité eN Ile de France). The data on maximum bus loads has been established and consolidated by Ile-de-France Mobilités from information provided by its partners.

	Number of vehicles	Maximum bus load
Standard bus	94	99 passengers
Articulated bus	12	156 passengers

Table 5.15 – Details of the available bus fleet

5.3 Setting the bus filling rate parameter

We will now introduce the term “bus load” to describe the maximum load observed on a line trip. Table 5.16 presents an analysis of bus loads based on data obtained from a survey conducted by the operator. The survey includes observations of the maximum number of passengers observed in buses during a set of runs for all operated lines. In this context, a “bus run” refers to a single trip made by a bus from one terminus to another, during which passengers board and alight at various stops along the route. Based on the bus fleet parameters and time periods defined in Subsection 5.2, we propose to define the bus filling parameter based on the time period with the highest passenger demand per hour. For Monday to Friday operating days, this time period is from 16 : 30 to 18 : 30.

		Min. bus	Average bus	Median bus	Max. bus
		load obs.	load obs.	load obs.	load obs.
Monday-Friday	6h-8h	1	21.97	20.67	63
	8h-12h	1	17.57	15.6	56
	12h-14h	1.5	16.96	16.33	37.3
	14h-16h30	1.5	21.29	21	61
	16h30-18h30	1	22.79	20.33	49.5
	18h30-21h	1	13.03	8.67	47.5

Table 5.16 – Analysis of bus loads based on survey data collected from the operator

Table 5.17 shows the bus loads obtained from the FSP with the currently operated topology, using the specified bus filling rate parameter. To make the results as close as possible to the operator’s observations, we require that the bus lines be operated with a frequency difference of no more than 2 in each direction.

		Max. bus filling	Min. bus	Average bus	Median bus	Max. bus
		rate chosen	load obs.	load obs.	load obs.	load obs.
Monday-Friday 16 : 30 – 18 : 30	21%	12.74	21.06	20.53	32.76	
	22%	13.03	22.05	21.61	34.32	
	23%	13.78	23.45	22.69	35.88	
	24%	13.16	24.23	22.89	37.44	
	25%	12.91	25.14	23.61	39.00	

Table 5.17 – Bus loads obtained by solving the FSP with the chosen maximum bus filling rate parameter for all bus lines

Based on values in Tables 5.16 and 5.17, we propose to set a common value for the maximum bus filling rate parameter ϕ for all bus lines. We choose to set this maximum bus load parameter at 23% for the Monday through Friday periods.

5.4 Setting reference travel time deviation for passengers

The FSP model requires adherence to a service-level parameter called the *maximum deviation time of passengers from their reference travel time* (α^{max}). The reference travel time (ξ_{st}) for a passenger traveling from s to t is defined as the shortest possible path

duration on the network currently offered by the operator, with the frequencies operated, which includes the sum of the riding time and the expected waiting time. In our experiments, we set the (α^{max}) parameter to 100%, meaning that all passengers must use a path with a duration at most equal to the one currently offered by the operator with the current topology and frequencies operated.

However, for some passengers, it may be impossible to offer a reference time on the new topology as it was on the old topology. Therefore, we introduce a preprocessing step to identify passengers who can be offered a path that respects the α^{max} parameter set to 100%. The path on the old topology associated with the reference travel time is called p' , where t_p denotes the duration of the path p' , which equals $r_p + w_p$, where r_p is the in-vehicle travel time and w_p is the total waiting times with the frequencies operated.

We note $L_{p'}$ as the number of lines taken when using the path p' . For passengers to have a better journey on the new topology, their path p on the new topology must meet two conditions: (1) the in-vehicle travel time of path p must be at most equal to the in-vehicle travel time of the path p' on the old topology; and (2) the number of transfers associated with path p must be at most equal to the one of path p' on the old topology. These two conditions can be summarized as

- $r_p \leq r_{p'}$
- $|L_p| \leq |L_{p'}|$

Passengers who do not meet these conditions are exempted from complying with the α^{max} service level parameter.

5.5 Setting up the experimental framework

In the context of using PPSP on the Poitiers topologies, we identified certain areas that required adjustment. Subsequently, this subsection focuses on establishing the experimental framework.

5.5.1 Two hypothesis on passengers behavior

We apply PPSP to the pre- and post-restructuring networks as well as to the A , B , and C networks, using two passenger choice hypothesis: “directness” and “shortest path”. In accordance with the “shortest path” hypothesis, we assume that passengers select the path with the shortest duration as their preferred path. Therefore, we include all passenger paths with zero to two transfers in the ω set. For the “directness” hypothesis, we assume

that passengers choose the path with the minimum number of transfers with respect to the α^{max} service level parameter as described in Subsection 5.4. We stop adding paths with more transfers to the ω set once we find a path p that meets the α^{max} service level parameter, even if there are other paths with more transfers that may result in a shorter travel time. For passengers who are exempted from complying with the α^{max} service level parameter, we assume the “shortest path” hypothesis.

5.5.2 Limitations of frequency difference in both directions of a line

After running our initial PPSP experiments, we found it necessary to make adjustments to the line frequencies in order to improve the realism of our results. We noticed that some lines had very different frequencies in each direction, leading to a significant underestimation of the costs that the operator would incur when implementing these frequencies. For example, if a line has high frequency in one direction and low frequency in the other direction, the operator may have to run empty buses in the direction with low frequency, leading to increased costs without generating corresponding revenue. Additionally, passengers may be discouraged from using the bus line if they perceive it as unreliable due to the different frequencies in each direction, resulting in reduced ridership. To address this issue, based on the frequencies currently used by the operator, we decided to limit the maximum frequency difference to 2 between the two directions of a line by adding a constraint in the model. We adopted this assumption for our subsequent experiments.

5.5.3 Bus traffic restriction

In consultation with the operator, we have decided to impose certain restrictions on the use of articulated buses on certain lines. For the pre-restructuring network, we have observed that the operator uses articulated buses on lines 1, 1 (east), and 2. Based on this, we prevent the use of articulated buses on lines other than these three lines. After restructuring, articulated buses are only used on lines 1 and 1 (east). Therefore, to ensure the operational feasibility of using articulated buses, we prohibit lines other than 1, 1 (east) from using articulated buses when determining frequencies on the post-restructuring network as well as on networks A , B and C .

5.6 Setting the frequencies

Table 5.18 – PPSP frequencies for business days obtained for the operator networks and for the proposed networks *A*, *B* and *C*

		Direct path hypothesis			Shortest path hypothesis		
		Est. kms costs	% direct pass.	Avg pass. travel time	Est. kms costs	% direct pass.	Avg pass. travel time
Pre-restruct netw.	6h-8h	8246 €	70.39%	22.43 min	8336 €	67.09%	22.32 min
	8h-12h	13271 €	74.01%	20.63 min	13358 €	70.57%	20.50 min
	12h-14h	6175 €	74.89%	21.73 min	6265 €	71.35%	21.69 min
	14h-16h30	8647 €	74.45%	20.12 min	8693 €	70.87%	19.74 min
	16h30-18h30	8451 €	73.49%	22.87 min	8277 €	69.81%	22.76 min
	18h30-21h	3751 €	73.45%	19.89 min	3770 €	69.57%	19.77 min
Post-restruct netw.	6h-8h	6997 €	74.35%	22.23 min	7433 €	69.46%	22.16 min
	8h-12h	8938 €	76.99%	20.64 min	9075 €	72.88%	20.55 min
	12h-14h	5574 €	77.90%	21.72 min	5829 €	72.53%	21.61 min
	14h-16h30	6439 €	75.80%	20.29 min	6362 €	73.26%	20.19 min
	16h30-18h30	7520 €	75.98%	22.91 min	7626 €	71.31%	22.79 min
	18h30-21h	3085 €	77.47%	19.79 min	3088 €	72.64%	19.71 min
Prop. network <i>A</i>	6h-8h	7303 €	77.12%	21.80 min	7473 €	72.76%	21.64 min
	8h-12h	9985 €	80.62%	20.18 min	9782 €	76.98%	20.07 min
	12h-14h	6037 €	82.23%	21.25 min	6037 €	78.92%	21.10 min
	14h-16h30	7316 €	79.73%	19.85 min	7126 €	76.34%	19.73 min
	16h30-18h30	8790 €	78.42%	22.40 min	8512 €	74.19%	22.24 min
	18h30-21h	3707 €	82.55%	19.32 min	3618 €	79.39%	19.21 min
Prop. network <i>B</i>	6h-8h	6649 €	81.91%	21.36 min	6660 €	78.87%	21.23 min
	8h-12h	8958 €	85.72%	19.79 min	8618 €	83.23%	19.68 min
	12h-14h	5211 €	85.87%	20.83 min	5214 €	84.14%	20.72 min
	14h-16h30	6731 €	84.56%	19.45 min	6574 €	82.67%	19.34 min
	16h30-18h30	7816 €	83.12%	22.01 min	7853 €	80.56%	21.87 min
	18h30-21h	3174 €	86.63%	18.98 min	3124 €	83.91%	18.83 min
Prop. network <i>C</i>	6h-8h	7480 €	79.09%	21.40 min	7332 €	75.60%	21.29 min
	8h-12h	9945 €	83.94%	19.80 min	9775 €	80.55%	19.70 min
	12h-14h	6100 €	85.19%	20.88 min	6118 €	81.94%	20.76 min
	14h-16h30	7543 €	83.49%	19.48 min	7419 €	80.36%	19.36 min
	16h30-18h30	9520 €	81.58%	22.03 min	9518 €	77.99%	21.89 min
	18h30-21h	4119 €	85.47%	19 min	3940 €	80.53%	18.97 min

Table 5.18 describes the frequencies obtained by solving PPSP on Monday-Fridays for the operator pre- and post-restructuring networks and for the three proposed networks *A*, *B* and *C* obtained with CGEP. From Table 5.18 we observe a significant improvement for each time period in the post-restructuring network of the operator, which has a much better potential in terms of potential kilometric costs and service-levels compared to their pre-restructuring network. We also observe that the use of PPSP to determine frequencies for the proposed networks *A*, *B*, and *C* results in a higher percentage of direct travel and more comfortable travel times for passengers than the post-restructuring network with frequencies set by PPSP. Finally, we observe that the proposed networks *A*, *B*, and *C* maintain the percentage of direct travelers specified in the design stage for every time period, ensuring a consistently high level of direct travel throughout the day.

Table 5.19 – Comparison of network operating costs obtained with PPSP for Monday-Fridays

	Daily est. kms costs	Daily adj. kms costs	Avg % dir. pass.	Avg mean pass. travel time
Pre-restr. + PPSP directness	48541 €	54851 €	73.52%	21.37 min
Pre-restr. + PPSP shortest path	48699 €	55030 €	69.98%	21.21 min
Pre-restr. + PPSP mix dir./SP	48547 €	54858 €	72.05%	21.32 min
Post-restr. + PPSP directness	38553 €	43565 €	76.32%	21.37 min
Post-restr. + PPSP shortest path	39413 €	44537 €	72.04%	21.28 min
Post-restr. + PPSP mix dir./SP	39350 €	44465 €	73.87%	21.32 min
Network A + PPSP directness	43138 €	48746 €	79.83%	20.91 min
Network A + PPSP shortest path	42548 €	48079 €	76.08%	20.77 min
Network A + PPSP mix dir./SP	43030 €	48624 €	77.83%	20.83 min
Network B + PPSP directness	38539 €	43549 €	84.47%	20.51 min
Network B + PPSP shortest path	34919 €	39458 €	82.09%	20.39 min
Network B + PPSP mix dir./SP	38590 €	436007 €	83.24%	20.45 min
Network C + PPSP directness	44707 €	50519 €	82.91%	20.54 min
Network C + PPSP shortest path	44102 €	49835 €	79.43%	20.42 min
Network C + PPSP mix dir./SP	44575 €	50370 €	81.19%	20.47 min

Table 5.19 provides a summary of the kilometric costs and observed passenger service-levels for the Monday-Friday period for both the pre- and post-restructuring networks, as well as the *A*, *B*, and *C* networks generated using CGEP. The first column shows the

estimated kilometric costs obtained from the output of PPSP, while the second column provides the adjusted kilometric costs, which are calculated by multiplying the estimated kilometric costs with the adjustment coefficient described in Subsection 4.3.

The estimated operating costs are described by three modes: “PPSP directness”, “PPSP shortest path”, and “PPSP mix direct./SP”. For each of these modes, the reference travel times of passengers are computed based on the post-restructuring topology and with the frequencies offered during the 2020 – 2021 season. The “PPSP directness” and “PPSP shortest path” modes estimate the costs based on the two passenger hypothesis described in Subsection 5.5. The “PPSP mix direct./SP” mode involves using “passengers directness” hypothesis during off-peak hours and “passengers shortest-path” hypothesis during peak hours. The off-peak and peak hours are defined by the operator as 8:00-12:00, 14:00-16:30, and 18:30-21:00, and 6:00-8:00, 12:00-14:00, and 16:30-18:30, respectively. We use different hypothesis to reflect the different passenger demand during peak and off-peak hours on Monday-Friday periods. The direct passenger percentage and the mean passenger travel time are computed by weighting the values obtained for each period with the number of passengers who use the network during that period, as specified in Table 5.14.

From Tables 5.18 and 5.19, we observe a significant improvement in the potential for optimizing frequencies in the post-restructuring network compared to the pre-restructuring network. Therefore, the new line design appears to be of high quality, allowing for more cost savings by optimizing its frequencies than by optimizing the frequencies of the pre-restructuring network’s lines. However, it is important to keep in mind that these estimated kilometric costs may be challenged during the timetable development stage. In addition, the OD demands may also change over time and there could be variations from the ones observed during the 2013 survey.

In addition, we found that optimizing the frequencies of networks A and C resulted in more expensive line plans than optimizing the frequencies of the post-restructuring network of the local operator. Moreover, we observe that when using PPSP to determine the frequencies for the proposed networks A , B , and C , we find a high percentage of direct passengers that is consistent with the service-level goals set in CGEP during the design stage.

On the other hand, we found that optimizing the frequencies of network B led to an interesting and cost-efficient line plan, which appeared to be more economical than the line plan obtained by optimizing the frequencies of the post-restructuring network

of the local operator, in particular when considering “passengers directness” path choice. Moreover, the line plan obtained by applying PPSP to network B appears to offer a higher level of service quality compared to the line plan that can be obtained by applying PPSP to the post-restructuring network.

This conclusion is particularly interesting because Network A was created taking $\alpha = 1.75$ and the LPP-SL cost estimation for this network was lower than for network B with $\alpha = 1.5$. However, by introducing passenger choices in a bi-level formulation of the FSP, we find that a lower value of α at the network design level reduces the approximation from the real passenger flow estimation. Hence, when frequencies are adjusted in the FSP, a better cost estimation is provided and Network B is found to have a better cost.

5.7 Synthesis

Based on our experimentation, we found that the post-restructuring network is better suited to meet travel demand on weekday periods compared to the pre-restructuring network, due to its higher optimization potential for frequencies and lower operating costs. Therefore, optimizing the post-restructuring network can result in better frequency plans leading to cost savings for both periods. In conclusion, the post-restructuring network established by the local operator offers higher quality service-levels and lower operating costs than their pre-restructuring network.

Optimizing the frequencies of networks A and C resulted in more expensive line plans compared to the frequencies optimized for post-restructuring network. However, the service-level they offer is better than the ones of pre and post-restructuring networks.

Compared to the line plan obtained with PPSP for the post-restructuring network, network B has a slightly less expensive line plan with better service-level quality for passengers. This means that overall, network B offers the most cost-effective option while still providing better guaranteed service quality to passengers compared to the other proposed networks.

For more information on the CGEP-generated lines that compose bus network B, please refer to Appendix E, which provides a detailed summary. In addition, see Appendices F.3 and F.4 to view the line frequencies obtained with PPSP for the post-restructuring network and bus network B .

CONCLUSION

The work presented in this manuscript was carried out within the framework of a CIFRE thesis between the LS2N and Lumiplan. In this thesis, we addressed two bus network optimization problems: the line planning problem with service levels, and the frequency setting problem. These two models aim to help public transport authorities, either to re-optimize their network or simply as a comparison to their existing network.

Methodological contributions

In this thesis we proposed two processes for the planning of urban transportation networks. We first proposed a Column Generation and Enumeration Process (CGEP) to solve the Line Planning Problem and get a set of bus lines to be operated. This process is based on a cost-oriented model 1.3 from the literature reinforced by operating and service level constraints. The CGEP is composed of two steps. The first step consists of simultaneously generating a set of bus lines and passenger paths using a column generation method. The passenger paths are separated into two pools according to their directness. This line generation method differs from the majority of methods in the literature which generally use a pool of lines defined as input and then select lines from this pool. The mixed integer model is solved with a commercial solver and allows to give an upper bound to the problem. The second step of the process consists in enumerating all the bus lines and passenger paths having a reduced cost between the lower and upper bounds obtained in the first phase. For both the generation and enumeration of passenger paths and bus lines, two pricing algorithms have been proposed, based on a labeling method.

Second, we proposed a process to solve the Frequency Setting Problem and determine operating frequencies for different time periods for a set of bus lines. In this problem, a challenge is to take into account the path choices made by passengers. The proposed process is based on a single-level cost-oriented model 1.2 taking into account shortest path constraints for the passengers, including expected waiting times. This single-level reformulation is made possible by using a discrete set of line frequencies and boolean variables to assign frequencies from this set to bus lines. The number of passenger paths

that can be taken by passengers can quickly become enormous as the size of the graph and the number of bus lines operated increases. We therefore proposed a Passenger Path Selection Process (PPSP) in order to select the passenger paths that are interesting for the passengers and lead to reasonable operational costs for the operator.

Contributions in terms of application

A case study was defined on the agglomeration of Poitiers. To define this case study, we created a graph from the GTFS data of the operator and the OpenStreetMap tool. We determined the street sections taken by the lines in order to compose the arcs of our graph. The determination of the nodes of the graph was made possible by selecting a set of major stations of the network according to their number of boarding and alighting travelers as well as according to their geographical position within different IRIS zones.

The OD matrices were generated by cross-referencing the user trips between IRIS zones with the boarding and alighting data of each major station. We thus obtained a set of OD matrices linking the nodes of our graph and covering one day of operation. In addition to setting up the case study, we also developed a line typology that we integrated into our bus line pricing algorithm. The typology rules, based on the length and topology of the lines, were defined based on a literature review and the analysis of four networks from French cities.

The objective of the experimentation of our models on this case study was to answer a question asked by the transport authority of Poitiers: "Did we make judicious restructuring choices?" In addition to answering this question by judging the relevance of this restructuring, both from the point of view of costs and service quality, we also proposed additional improvements to the network. The application of CGEP to this case study to define a set of lines resulted in three possible network redesign options. These options maintain at least the same quality of service parameters as those proposed by the operator and offer potential gains in terms of operating costs. The determination of the frequencies was then made with PPSP, on the current network of Poitiers as well as on the three ones determined with CGEP. Using PPSP, we were able to compare the three options generated by CGEP and identify one as the most promising in terms of both operating costs and service quality.

Research perspectives

In our experiments on the line planning problem, we noticed that most algorithms require a functional network with pre-defined service-level parameters to start with. This means that the initial solution must already meet the desired service-level parameters, which can be challenging to find. Therefore, future research could focus on developing heuristics or optimization methods that can find an initial solution that meets the desired service-level parameters.

We also observed that the estimated passenger flows in the line planning stage may not match the actual flows that occur at the frequency setting stage. To address this issue, one possible solution could be to adapt the maximum-spt-deviation service-level parameter for each passenger based on their shortest in-vehicle travel time on the operated network. This approach could help improve the accuracy of the estimated passenger flows and ultimately lead to more effective frequency setting.

Distributing passengers across multiple paths for the same origin-destination (OD) pair during the frequency setting stage has been recognized as an important area of research since Spiess and Florian [1989]. While integrating the multiple distribution of passengers across paths into a bilevel model presents challenges, it can provide a more comprehensive and realistic representation of passenger flows in the transit network.

Achieving a smooth frequency transition between different periods is a key factor to consider when designing timetables for transit networks. Abrupt changes in frequency can result in operational difficulties and dissatisfaction among passengers. Integrating this frequency smoothing during the frequency setting stage could help prevent operational challenges and passenger dissatisfaction resulting from sudden changes in frequency, and lead to a more gradual and realistic transition between different periods.

Finally, Lumiplan would greatly benefit from having a simulator that considers passenger travel and congestion to assess the overall quality of public transportation networks. Such a simulator would allow to judge the performance of the network as a whole, including the design of the lines and the frequencies assigned to them. It would also help to identify potential problems, such as bottlenecks and overcrowding, and test different scenarios to optimize the network's design. Therefore, the company may consider acquiring or developing such a tool to improve the accuracy of their network planning and analysis.

APPENDICES

A Line Planning Problem

A.1 Overestimation of the number of direct travelers with DC constraints

This subsection describes a small example concerning the overestimation of the number of direct travelers with DC constraints.

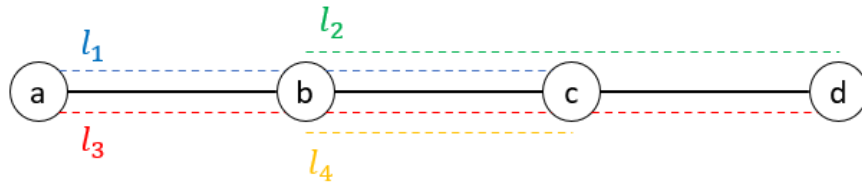


Figure 16 – A bus network with four nodes : $\{a, b, c, d\}$ and four bus lines : $l_1 = (a, b, c)$, $l_2 = (b, c, d)$, $l_3 = (a, b, c, d)$ and $l_4 = (b, c)$.

On Figure 16, suppose we have a (a, c) OD pair with 100 travelers, a (b, d) OD pair with 100 travelers and a (a, d) OD pair with 100 travelers. Furthermore, we assume the capacity of a bus to be set at 100 passengers.

By looking at Figure 16 we can see that travelers of the (a, c) OD pair can be served with a direct path through the arc $[b, c]$ by the lines l_1 and l_3 . Travelers of the (b, d) OD pairs can be served by the lines l_2 and l_3 , and travelers of the (a, d) OD pairs can be served by the line l_3 . Hence, we have :

- ★ For the arc (a, b) , $\mathcal{L}_{ac}(a, b) = \{l_1, l_3\}$, $\mathcal{L}_{bd}(a, b) = \emptyset$, $\mathcal{L}_{ad}(a, b) = \{l_3\}$.
- ★ For the arc (b, c) , $\mathcal{L}_{ac}(b, c) = \{l_1, l_3\}$, $\mathcal{L}_{bd}(b, c) = \{l_2, l_3\}$, $\mathcal{L}_{ad}(b, c) = \{l_3\}$.
- ★ For the arc (c, d) , $\mathcal{L}_{ac}(c, d) = \emptyset$, $\mathcal{L}_{bd}(c, d) = \{l_2, l_3\}$, $\mathcal{L}_{ad}(c, d) = \{l_3\}$.

The instinctive choice consists to set a frequency of 1 to the lines l_1 , l_2 and l_3 and a frequency of 0 to the line l_4 . However, setting l_3 to 2 and l_4 to 1 suffices to satisfy the

DC constraint even if the (a, c) and the (b, d) travelers cannot use this path without to transfer in reality. Indeed, the equivalence classes are defined as follows:

- For the arc (a, b) : (a, c) and (a, d) OD pairs have at least one line from their origin to their destination through it. (a, c) and (a, d) do not have the same allowed lines pool, thus we have two equivalence classes, each of them composed of one of the two OD pairs.
- For the arc (b, c) : (a, c) , (b, d) and (a, d) OD pairs have at least one line from their origin to their destination through it. (a, c) , (b, d) and (a, d) do not have the same allowed lines pool, thus we have three equivalence classes, each of them composed of one of the three OD pairs.
- For the arc (c, d) : (a, d) and (b, d) OD pairs have at least one line from their origin to their destination through it. (a, d) and (b, d) do not have the same allowed lines pool, thus we have three equivalence classes, each of them composed of one of the three OD pairs.

Thus, we obtain 7 different DC constraints, two for the arc (a, b) , three for the arc (b, c) and two for the arc (c, d) . Suppose the $((a, b), (b, c))$, $((a, b), (b, c), (c, d))$ and $((b, c), (c, d))$ rdc-passenger-paths are known and present in the rdc-paths pool \mathcal{P}^{0+} . Furthermore, suppose all these passenger paths respect the maximum-spt-deviation. These 7 constraints are as follows :

- *For the arc (a, b) and the (a, c) equivalence class* : both (a, c) and (a, d) OD pairs $\in [a, c]_{(a,b)}^{\leq, \alpha}$ as $L_{ac}(a, b) \subseteq L_{ac}(a, b)$ and $L_{ad}(a, b) \subseteq L_{ac}(a, b)$.
Thus we have $y_{((a,b),(b,c))}^{0+} + y_{((a,b),(b,c),(c,d))}^{0+} \leq 100 * f_{l_1} + 100 * f_{l_3}$.
- *For the arc (a, b) and the (a, d) equivalence class* : (a, d) OD pair $\in [a, d]_{(a,b)}^{\leq, \alpha}$ as $L_{ad}(a, b) \subseteq L_{ad}(a, b)$.
Thus we have $y_{((a,b),(b,c),(c,d))}^{0+} \leq 100 * f_{l_3}$.
- *For the arc (b, c) and the (a, c) equivalence class* : both (a, c) and (a, d) OD pairs $\in [a, c]_{(b,c)}^{\leq, \alpha}$ as $L_{ac}(b, c) \subseteq L_{ac}(b, c)$ and $L_{ad}(b, c) \subseteq L_{ac}(b, c)$.
Thus we have $y_{((a,b),(b,c))}^{0+} + y_{((a,b),(b,c),(c,d))}^{0+} \leq 100 * f_{l_1} + 100 * f_{l_3}$
- *For the arc (b, c) and the (b, d) equivalence class* : both (a, d) and (b, d) OD pairs $\in [b, d]_{(b,c)}^{\leq, \alpha}$ as $L_{ad}(b, c) \subseteq L_{bd}(b, c)$ and $L_{bd}(b, c) \subseteq L_{bd}(b, c)$.
Thus we have $y_{((a,b),(b,c),(c,d))}^{0+} + y_{((b,c),(c,d))}^{0+} \leq 100 * freq\ l_2 + 100 * freq\ l_3$
- *For the arc (b, c) and the (a, d) equivalence class* : (a, d) OD pair $\in [a, d]_{(b,c)}^{\leq, \alpha}$ as $L_{ad}(b, c) \subseteq L_{ad}(b, c)$.

Thus we have $y_{((a,b),(b,c),(c,d))}^{0+} \leq 100 * f_{l_3}$

— For the arc (c, d) and the (b, d) equivalence class : both (a, d) and (b, d) OD pairs $\in [b, d]_{(c,d)}^{\leq, \alpha}$ as $L_{ad}(c, d) \subseteq L_{bd}(c, d)$ and $L_{bd}(c, d) \subseteq L_{bd}(c, d)$.

Thus we have $y_{((a,b),(b,c),(c,d))}^{0+} + y_{((b,c),(c,d))}^{0+} \leq 100 * f_{l_2} + 100 * f_{l_3}$

— For the arc (c, d) and the (a, d) equivalence class : (a, d) OD pair $\in [a, d]_{(c,d)}^{\leq, \alpha}$ as $L_{ad}(c, d) \subseteq L_{ad}(c, d)$

Thus we have $y_{((a,b),(b,c),(c,d))}^{0+} \leq 100 * f_{l_3}$

$y_{((a,b),(b,c))}^{0+}$, $y_{((a,b),(b,c),(c,d))}^{0+}$, and $y_{((b,c),(c,d))}^{0+}$ have to be set to 100 to satisfy the associated travelers. These 7 DC constraints can be respected by setting a frequency of 2 to the line l_3 and a frequency of 0 to the lines l_1 , l_2 and l_4 .

Going back to model 1.3, constraint (3.3) sets the lines frequencies according to the travelers on its respective arcs. As it is not respected on the arc (b, c) , a solution consists to put a frequency of 1 to the line l_4 . We finally get a frequency of 2 for the line l_3 , a frequency of 1 for the line l_4 and a frequency of 0 for lines l_1 and l_2 . Nevertheless, one of the three travelers demands has to transfer from the line l_3 to the line l_4 even if these three travelers demands are considered as direct in the model. This is the limit of this DC constraint.

A.2 LPP-SL linearization

Appendix describing the linearization of the LPP-SL using the variable substitution of Borndörfer et al. [2007]. The goal of this linearization is to model the use of a line and its frequency as a single variable.

$$f_{min} \times z_l \leq f_l \leq f_{max} \times z_l \quad \forall l \in \mathcal{L} \quad (1)$$

$$\iff \frac{f_l}{f_{max}} \leq z_l \leq \frac{f_l}{f_{min}} \quad \forall l \in \mathcal{L} \quad (2)$$

$$\implies \sum_{l \in \mathcal{L}} \frac{f_l}{f_{max}} \leq \sum_{l \in \mathcal{L}} z_l \leq \sum_{l \in \mathcal{L}} \frac{f_l}{f_{min}} \quad (3)$$

$$\iff f_{min} \times \sum_{l \in \mathcal{L}} z_l \leq \sum_{l \in \mathcal{L}} f_l \leq f_{max} \times \sum_{l \in \mathcal{L}} z_l \quad (4)$$

$$\implies \sum_{l \in \mathcal{L}} f_l \leq f_{max} \times n \quad (5)$$

A.3 LPP-SL master and dual problems

LPP-SL Master Problem

$$\min \lambda \times \sum_{l \in \mathcal{L}} f_l \times c \times d_l + (1 - \lambda) \times \left(\sum_{p \in \mathcal{P}^{0+}} y_p^{0+} \times r_p + \sum_{p \in \mathcal{P}^1} y_p^1 \times r_p \right) \quad (6)$$

$$\text{s.t.} \quad \sum_{p \in \mathcal{P}_{st}^{0+}} y_p^{0+} + \sum_{p \in \mathcal{P}_{st}^1} y_p^1 = q_{st} \quad \forall (s, t) \in D \quad (3.2) \quad (\pi_{st})$$

$$\sum_{p \in \mathcal{P}^{0+} \mid a \in p} y_p^{0+} + \sum_{p \in \mathcal{P}^1 \mid a \in p} y_p^1 \leq \sum_{l \in \mathcal{L} \mid a \in l} \kappa \times f_l \quad \forall a \in A \quad (3.3) \quad (\mu_a)$$

$$\sum_{(u,v) \in [s,t]_{\tilde{a}}^{\leq, \alpha}} \sum_{p \in \mathcal{P}_{uv}^{0+}(a)} y_p^{0+} \leq \sum_{l \in \mathcal{L}_{st}^{\alpha}(a)} \kappa \times f_l \quad \forall a \in A, \forall [s, t]_a \in \tilde{D}(a) \quad (3.11) \quad (\nu_{a, [s, t]_a})$$

$$\sum_{l \in \mathcal{L} \mid a \in l} f_l \leq s_a \quad \forall a \in A \quad (3.5) \quad (\eta_a)$$

$$\sum_{p \in \mathcal{P}^{0+}} y_p^{0+} \geq \sum_{(s,t) \in D} \beta \times q_{st} \quad (3.6) \quad (\psi_1)$$

$$\sum_{l \in \mathcal{L}} 2 \times r_l \times f_l \leq o \times \Delta \quad (3.7) \quad (\theta)$$

$$\sum_{l \in \mathcal{L}} f_l \leq f_{max} \times n \quad (5) \quad (\omega)$$

$$y_p^{0+} \in \mathbb{R}_+ \quad \forall p \in \mathcal{P}^{0+}$$

$$y_p^1 \in \mathbb{R}_+ \quad \forall p \in \mathcal{P}^1$$

$$f_l \in [0, f_{max}] \quad \forall l \in \mathcal{L}$$

LPP-SL Dual Problem

$$\begin{aligned} \max \quad & \sum_{(s,t) \in D} \pi_{st} \times q_{st} - \sum_{a \in A} \eta_a \times s_a + \psi_1 \times \sum_{(s,t) \in D} \beta \times q_{st} \\ & - \theta \times o \times \Delta - \omega \times f_{max} \times n \end{aligned} \quad (7)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{a \in l} \mu_a \times \kappa + \sum_{a \in l} \sum_{[s,t]_a \in \tilde{D}(a) \mid l \in \mathcal{L}_{st}^\alpha(a)} \nu_{a,[s,t]_a} \times \kappa \\ & - \sum_{a \in l} \eta_a - 2 \times r_l \times \theta - \omega \leq \lambda \times c \times d_l \quad \forall l \in \mathcal{L} \end{aligned} \quad (8)$$

$$\begin{aligned} \pi_{uv} - \sum_{a \in \mathcal{P}} \mu_a \\ - \sum_{a \in \mathcal{P}} \sum_{[s,t]_a \in \tilde{D}(a) \mid (u,v) \in [s,t]_a^{\leq, \alpha}} \nu_{a,[s,t]_a} + \psi_1 \leq (1 - \lambda) \times r_p \quad \forall (u,v) \in D, \forall p \in \mathcal{P}_{uv}^{0+} \end{aligned} \quad (9)$$

$$\pi_{uv} - \sum_{a \in \mathcal{P}} \mu_a \leq (1 - \lambda) \times r_p \quad \forall (u,v) \in D, \forall p \in \mathcal{P}_{uv}^1 \quad (10)$$

$$\pi_{st} \quad \text{unrestricted} \quad \forall (s,t) \in D \quad (11)$$

$$\mu_a, \eta_a \geq 0 \quad \forall a \in A \quad (12)$$

$$\nu_{a,[s,t]_a} \geq 0 \quad \forall a \in A, \forall [s,t]_a \in D_a \quad (13)$$

$$\psi_1, \theta, \omega \geq 0 \quad (14)$$

A.4 Bus Line Labeling Algorithm

As a reminder, a bus line label ϵ is defined by the tuple (v, c, δ, N, R, S) where v denotes the last node visited, c the reduced cost of the label, δ the distance of label, N the set of nodes visited by the label and R the set of OD pairs satisfied by the label because they have a direct path on it and S the set of OD pairs not directly satisfied by the label whatever its future extension.

Algorithm 1 Function to generate lines from a source to a destination

```
1: function GenerateLines(src: source, dst: destination, g: graph,  $d_{max}$ : maximum
   Bus Line distance allowed, duals: duals values of the constraints of the master problem)

2: L: List of line labels
3: E: List of line labels
4:  $L := \{(src, 0, 0, \{src\}, \{\}, \{\})\}$ 
5:  $E := \emptyset$ 
6: while  $\sum_{\epsilon \in L: v_\epsilon \neq dst} |L| \neq 0$  do
7:   pick a label  $\epsilon$  in L with  $v_\epsilon \neq dst$ 
8:    $L \leftarrow L \setminus \epsilon$ 
9:    $E \leftarrow E \cup \epsilon$ 
10:  for all nodes  $j \in V \setminus N_\epsilon$ , linked to  $v_\epsilon$  by an arc in the graph G do
11:    if  $\delta_\epsilon + d_{(v_\epsilon, j)} + d_{(j, v_\epsilon)} \leq d_{max}$  then
12:      Extension of the label  $\epsilon$  along arc  $(v_\epsilon, j)$  with duals values to create label  $\epsilon'$ 
13:      for all equivalence classes  $[s, t]_a$  where  $a = (v_\epsilon, j)$  do
14:         $cost\_Line\_Update([s, t]_a, \epsilon, \epsilon', duals)$ 
15:      end for
16:      for all OD pairs  $(s, t) \notin R_{\epsilon'} \cup S_{\epsilon'}$  do
17:         $R\_Update((s, t), \epsilon')$ 
18:      end for
19:      if  $\epsilon'$  is not dominated by a label in  $L_j \cup E_j$  then
20:         $L \leftarrow L \cup \epsilon'$ 
21:        Removing all labels dominated by  $\epsilon'$  from  $L_j \cup E_j$ 
22:      end if
23:    end if
24:  end for
25: end while
26: return all labels  $\epsilon$  in  $L_{dst}$  with  $c_\epsilon < 0$ 
27: end function
```

Algorithm 2 Function that extends the reduced cost of the label according to the equivalence classes that have a direct path by using the new label ϵ'

- 1: function **cost_Line_Update**(Equivalence class $[s, t]_a$, label ϵ , label ϵ' , *duals*: duals values of the constraints of the master problem)
 - 2: **if** (the OD pair $(s, t) \notin R_{\epsilon'} \cup S_{\epsilon'} \cap (s$ belongs to N_{ϵ} and t is equal to j , or the opposite) **then**
 - 3: **if** the duration of the path from s to t on the created line respects the maximum-spt-deviation **then**
 - 4: Update of the reduced cost $c_{\epsilon'}$ on all the arcs relying the nodes of the label ϵ' between u and w : $c_{\epsilon'} \leftarrow c_{\epsilon'} - \sum_{a \in A_{N_{\epsilon'}, (u, w)}} (\nu_{a, (u, w)} + \nu_{\bar{a}, (u, w)}) \times \kappa$.
 - 5: Add all the OD pairs being dc-equivalent to (s, t) wrt a to the $S_{\epsilon'}$ set
 - 6: **end if**
 - 7: **end if**
-

Algorithm 3 Function that determines the OD pairs that can no longer have a direct path by using the new label ϵ'

- 1: function **R_Update**(OD pair (s, t) , label ϵ')
 - 2: **if** the duration of the path from s to t on the created line exceeds the maximum-spt-deviation **then**
 - 3: Add the OD pair (s, t) to the $R_{\epsilon'}$ set
 - 4: **end if**
-

Examples A.1 and A.2 show how the **cost_Update** and **R_Update** functions work on a miniature network.

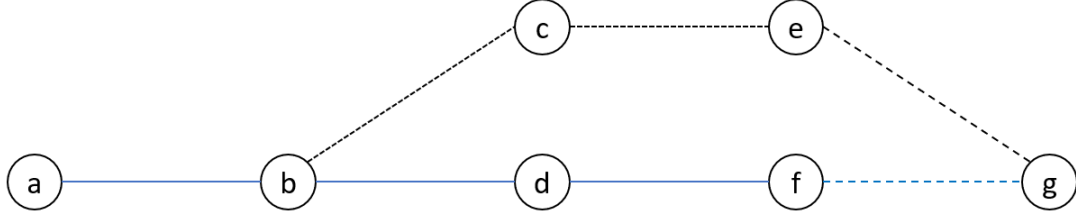


Figure 17 – A small network represented with a graph composed of 7 nodes.

Example A.1. Example of cost_Line_Update execution

On Figure 17, suppose we are in presence of a label ϵ with $N_\epsilon = a, b, d, f$ and being extended to node g among the arc (f, g) . All equivalence classes defined on the arc (f, g) are then examined to identify if their origin is in the set N_{ϵ} and their destination is the node lastly added to extend the label ϵ or the opposite. Suppose (b, g) equivalence class exists with respect to the arc (f, g) . $[b, g]_a$ respects the conditions described above as $b \in N_{\epsilon}$ and g is the lastly added node. We check if the duration of the path from b to g on the created line respects the maximum-spt-deviation, if it does, then the passengers of the (b, g) OD pair have a direct passenger path by using the line associated to this label. The reduced cost $c_{\epsilon'}$ is then updated according to the dual values of constraints (3.11) on the arcs of the created line between b and g . Finally, the OD pairs that are dc-equivalent to (b, g) with respect to the arc (f, g) , and thus belong to the equivalence class $[s, t]_a$, are added to the set $S_{\epsilon'}$.

Example A.2. Example of R_Update execution

On Figure 17, let ϵ be a label with $N_\epsilon = a, b, d, f$. Let us assume that ϵ is extended to the node g along arc (f, g) . The new created label is called ϵ' . Suppose there exists a (b, e) OD pair and that this OD pair is not present in $R_{\epsilon'}$ nor in $S_{\epsilon'}$. Let us call $r_{(\epsilon', bg)}$ the duration of the path from b to g on the created line and $r_{s,t}^{min}$ the minimum in-vehicle travel time to go from s to t on the graph considered. If the ratio between $r_{(\epsilon', bg)} + r_{g,e}^{min}$ and t_{be}^{min} exceeds the maximum-spt-deviation enforced by the service level parameter α , the (b, e) OD pair can no longer have a direct passenger path on the label, regardless of future extensions to other nodes. This OD pair is then added to the set $R_{\epsilon'}$.

A.5 Passenger Path Labeling Algorithm

As a reminder, a passenger path label is defined by the tuple (v, c, τ, N) where v represents the last node visited by the label, c the reduced cost of the label, τ the time of the label and N the set of nodes visited by the label.

Algorithm 4 Function for generating attractive paths for a passenger from its source node to its destination node

```

1: function GeneratePath(src: source, dst: destination, G: graph, duals: duals values
   of the constraints of the master problem, pathDesired: type of path desired to be
   generated,  $\mathcal{L}$ : line pool)
2:  $L$ : List of path labels
3:  $E$ : List of path labels
4: if pathDesired = rdc then
5:    $L := \{(src, -\pi_{(src,dst)} - \psi_1, 0, \{src\}, \{\})\}$ 
6: else if pathDesired = indirect then
7:    $L := \{(src, -\pi_{(src,dst)}, 0, \{src\}, \{\})\}$ 
8: end if
    $E := \emptyset$ 
9: while  $\sum_{\epsilon \in L: v_\epsilon \neq dst} |L| \neq 0$  do
10:  pick a label  $E$  in  $L$  with  $v_\epsilon \neq dst$ 
11:   $L \leftarrow L \setminus \epsilon$ 
12:   $E \leftarrow E \cup \epsilon$ 
13:  for all nodes  $j$  of the graph  $G$  such that  $j \notin N_\epsilon$  and the arc  $(v_\epsilon, j)$  is an arc of  $G$ 
    do
14:     $\epsilon' = path\_Extension(\epsilon, src, dst, G, duals, pathDesired, \mathcal{L})$ 
15:  end for
16:  if  $\epsilon'$  is not dominated by a label in  $L_j \cup E_j$  then
17:     $L \leftarrow L \cup \epsilon'$ 
18:    Removing all labels dominated by  $\epsilon'$  from  $L_j \cup E_j$ 
19:  end if
20: end while
21: return all labels  $\epsilon$  in  $L_{dst}$  with  $c_\epsilon < 0$ 
22: end function

```

Algorithm 5 Function to extend the label of a passenger path

```
1: function path_Extension( $\epsilon$  : label to extend,  $src$ : source,  $dst$ : destination,  $G$ :  
   graph,  $duals$ : duals values of the constraints of the master problem,  $pathDesired$ :  
   type of path desired to be generated,  $\mathcal{L}$ : line pool)  
2: if  $\tau_\epsilon + r_{v_\epsilon, j} + r_{j, dst}^{min} \leq \alpha \times r_{src, dst}^{min}$  then  
3:   if (pathDesired = rdc and  $\exists l \in \mathcal{L}$  to go from  $src$  to  $j$  wrt the maximum-spt-  
     deviation) or (pathDesired = indirect and  $\exists l \in \mathcal{L}$  circulating on the arc  $(v_\epsilon, j)$ )  
     then  
4:     Extension of the label  $\epsilon$  along arc  $(v_\epsilon, j)$  to create label  $\epsilon'$   
5:     if pathDesired = rdc-passenger-path then  
6:        $cost\_rdcPath\_Update(\epsilon', (v_\epsilon, j), src, dst, duals)$   
7:     end if  
8:   end if  
9: end if  
10: return label  $\epsilon'$ 
```

Algorithm 6 Function that extends the reduced cost of the label in case of a rec-passenger path researched

```
1: function cost_rdcPath_Update(label  $\epsilon'$ , arc  $a$  extending the label  $\epsilon$ ,  $src$ : source,  
    $dst$ : destination,  $duals$ : duals values of the constraints of the master problem)  
2: for all equivalence classes  $[s, t]_a$  do  
3:   if  $(src, dst) \in [s, t]_a^{\leq, \alpha}$  then  
4:      $c_{\epsilon'} \leftarrow c_{\epsilon'} + \nu_{a, [s, t]_a}$   
5:   end if  
6: end for
```

Example A.3 shows how the **cost_Path_Update** function works on a miniature network.

Example A.3. Example of cost_rdcPath_Update execution

On Figure 17, we consider a label visiting nodes (a, b, d, f) and being extended to node g . Whether the passenger path we try to generate is a rdc or an indirect-passenger-path, it must be checked if the time of the label newly extended to the node g respects the service level α , meaning has an in-vehicle travel time smaller or equal than $\alpha \times$ the minimum in-vehicle travel time from a to g noted $r_{a, g}^{min}$.

Furthermore, in the case of a rdc-passenger-path generation from a to e , it must be checked if there is at least one line in the bus lines pool \mathcal{L} that allows to go from a to e with the arc (f, g) in it and by respecting the maximum-spt-deviation enforced by the service level parameter α . In the case of an indirect-passenger-path generation from a to e , it must be checked if there is at least one line in the bus lines pool \mathcal{L} that drives on the newly added arc (f, g) .

B Frequency Setting Problem

B.1 Passenger path requiring to be added to make the solution feasible

This subsection describes an example to illustrate the execution of the feasibility step introduced in 2.2. Figure 18 presents a small network in which a passenger path must enter the path base \mathcal{P} .

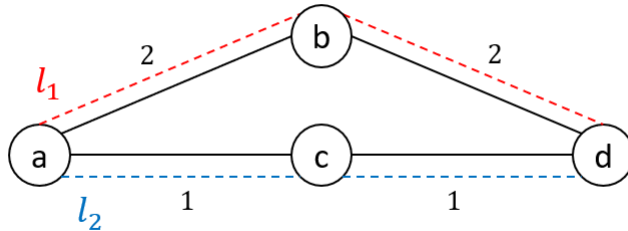


Figure 18 – Small network composed of four nodes and two bus lines $l_1 = (a, b, d)$ and $l_2 = (a, c, d)$.

Suppose first that we have an (a, d) OD pair and that the passenger path $p_1 = [(a, b), l_1], ((b, d), l_1)]$ belongs to the sets \mathcal{P} and Ω . In addition, suppose the existence of a passenger path $p_2 = [((a, c), l_2), ((c, d), l_2)]$ belonging to Ω but not to \mathcal{P} . Suppose that the frequencies of the two bus lines l_1 and l_2 are identical, which means the estimated waiting times on these lines are identical. Then, the passenger path p_2 is the shortest path from a to d on this network. As a result, the passenger path p_2 is the optimal solution of the follower and must be present in \mathcal{P} for the solution of $FSP(\mathcal{P})$ to be feasible.

C Literature instances

We use two literature instances, Dutch and Sioux-Falls. The Dutch instance is frequently used due to its relatively large size and complexity, consisting of a directed graph with 23 nodes and 60 arcs, along with an associated OD matrix of 420 non-zero OD pairs. The Sioux-Falls instance is also a commonly used benchmark in transportation studies, with a directed graph of 24 nodes and 76 arcs, and an associated OD matrix of 528 non-zero OD pairs. The relevant details about these two instances are listed in Table 20.

Table 20 – Details of the two instances considered

Instance	Nb of nodes	Nb of arcs	Nb of non-zero OD pairs	Total nb of passengers
Dutch	23	60	420	183 582
Sioux-Falls	24	76	528	360 600

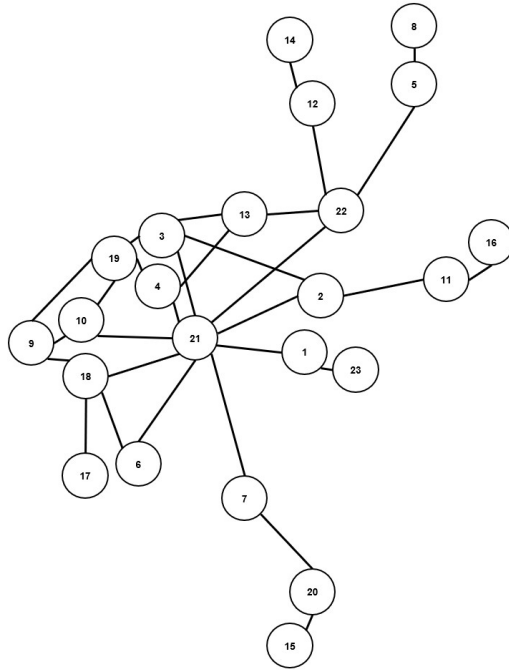
C.1 Dutch instance

Dutch instance details can be found in https://www.gams.com/latest/gamslib_ml/libhtml/gamslib_lop.html.

1	Ah	Arnhem
2	Apd	Apeldoorn
3	Asd	Amsterdam CS
4	Asdz	Amsterdam Zuid WTC
5	Asn	Assen
6	Bd	Breda
7	Ehv	Eindhoven
8	Gn	Groningen
9	Gv	Den Haag HS
10	Gvc	Den Haag CS
11	Hgl	Hengelo
12	Hr	Heerenveen
13	Lls	Lelystad Centrum
14	Lw	Leeuwarden
15	Mt	Maastricht
16	Odzg	Oldenzaal Grens
17	Rsdg	Rosendaal Grens
18	Rtd	Rotterdam CS
19	Shl	Schiphol
20	Std	Sittard
21	Ut	Utrecht CS
22	Zl	Zwolle
23	Zvg	Zevenaar Grens

Table 21 – Numerotation of the stations of the Dutch instance

Table 21 provides the names of the main stations associated with the nodes in the Dutch instance graph.



(a) The Dutch graph

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	0	0	726	721	16	136	101	0	109	741	0	11	8	13	0	0	320	180	60	0	4244	56	602
2	0	0	917	76	0	57	62	0	21	202	468	0	0	0	5	71	47	143	32	10	1160	0	83
3	726	917	0	0	235	154	437	205	730	2540	287	115	2783	90	37	22	2258	1756	6469	155	4919	400	489
4	721	76	0	0	58	138	542	88	461	207	24	16	819	21	149	0	6	369	1456	203	5826	171	155
5	16	0	235	58	0	42	33	1720	13	117	0	0	152	0	14	0	48	125	32	28	502	854	19
6	136	57	154	138	42	0	950	34	531	228	39	16	6	14	79	5	329	1829	7	157	1165	79	14
7	101	62	437	542	33	950	0	28	35	335	28	11	8	9	404	3	75	569	99	936	3109	47	11
8	0	0	205	88	1720	34	28	0	12	73	0	0	200	0	13	0	33	75	48	29	331	720	14
9	109	21	730	461	13	531	35	12	0	785	20	10	29	6	8	0	890	4586	1339	22	225	33	0
10	741	202	2540	207	117	228	335	73	785	0	81	48	31	26	41	7	3	2829	1503	104	3138	163	229
11	0	468	287	24	0	39	28	0	20	81	0	0	0	0	12	75	24	52	11	20	422	0	0
12	11	0	115	16	0	16	11	0	10	48	0	0	77	478	4	0	10	58	20	8	209	511	19
13	8	0	2783	819	152	6	8	200	29	31	0	77	0	77	1	0	15	46	103	2	89	390	0
14	13	0	90	21	0	14	9	0	6	26	0	478	77	0	4	0	7	36	20	9	145	380	14
15	0	5	37	149	14	79	404	13	8	41	12	4	1	4	0	0	22	73	29	863	359	21	0
16	0	71	22	0	0	5	3	0	0	7	75	0	0	0	0	0	0	11	0	21	0	0	0
17	320	47	2258	6	48	329	75	33	890	3	24	10	15	7	22	0	0	1077	164	19	325	100	0
18	180	143	1756	369	125	1829	569	75	4586	2829	52	58	46	36	73	11	1077	0	509	179	2260	182	157
19	60	32	6469	1456	32	7	99	48	1339	1503	11	20	103	20	29	0	164	509	0	44	278	64	0
20	0	10	155	203	28	157	936	29	22	104	20	8	2	9	863	21	19	179	44	0	720	46	0
21	4244	1160	4919	5826	502	1165	3109	331	225	3138	422	209	89	145	359	0	325	2260	278	720	0	1112	996
22	56	0	400	171	854	79	47	720	33	163	0	511	390	380	21	0	100	182	64	46	1112	0	32
23	602	83	489	155	19	14	11	14	0	229	0	19	0	14	0	0	0	157	0	0	996	32	0

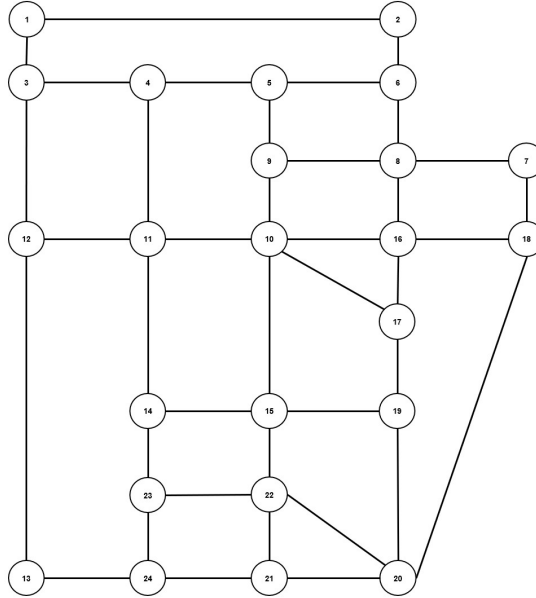
(b) The symmetrical OD matrix associated to the Dutch graph

Figure 19 – Dutch network and associated OD matrix

Similar to Borndörfer and Karbstein [2012] and Karbstein [2013], we consider that each edge has a length of one kilometer.

C.2 Sioux-Falls instance

The Sioux-Falls instance is described in Bar-Gera [2021].



(a) The Sioux-Falls graph

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0	100	100	500	200	300	500	800	500	1300	500	200	500	300	500	500	400	100	300	300	100	400	300	100
2	100	0	100	200	100	400	200	400	200	600	200	100	300	100	100	400	200	0	100	100	0	100	0	0
3	100	100	0	200	100	300	100	200	100	300	300	200	100	100	100	200	100	0	0	0	100	100	0	0
4	500	200	200	0	500	400	400	700	700	1200	1400	600	600	500	500	800	500	100	200	300	200	400	500	200
5	200	100	100	500	0	200	200	500	800	1000	500	200	200	100	200	500	200	0	100	100	100	200	100	0
6	300	400	300	400	200	0	400	800	400	800	400	200	200	100	200	900	500	100	200	300	100	200	100	100
7	500	200	100	400	200	400	0	1000	600	1900	500	700	400	200	500	1400	1000	200	400	500	200	500	200	100
8	800	400	200	700	500	800	1000	0	800	1600	800	600	600	400	600	2200	1400	300	700	900	400	500	300	200
9	500	200	100	700	800	400	600	800	0	2800	1400	600	600	600	900	1400	900	200	400	600	300	700	500	200
10	1300	600	300	1200	1000	800	1900	1600	2800	0	4000	2000	1900	2100	4000	4400	3900	700	1800	2500	1200	2600	1800	800
11	500	200	300	1500	500	400	500	800	1400	3900	0	1400	1000	1600	1400	1400	1000	100	400	600	400	1100	1300	600
12	200	100	200	600	200	200	700	600	600	2000	1400	0	1300	700	700	700	600	200	300	400	300	700	700	500
13	500	300	100	600	200	200	400	600	600	1900	1000	1300	0	600	700	600	500	100	300	600	600	1300	800	800
14	300	100	100	500	100	100	200	400	600	2100	1600	700	600	0	1300	700	700	100	300	500	400	1200	1100	400
15	500	100	100	500	200	200	500	600	1000	4000	1400	700	1300	0	1200	1500	200	800	1100	800	2600	1000	400	400
16	500	400	200	800	500	900	1400	2200	1400	4400	1400	700	600	700	1200	0	2800	500	1300	1600	600	1200	500	300
17	400	200	100	500	200	500	1000	1400	900	3900	1000	600	500	700	1500	2800	0	600	1700	1700	600	1700	600	300
18	100	0	0	100	0	100	200	300	200	700	200	200	100	100	200	500	600	0	300	400	100	300	100	0
19	300	100	0	200	100	200	400	700	400	1800	400	300	300	300	800	1300	1700	300	0	1200	400	1200	300	100
20	300	100	0	300	100	300	500	900	600	2500	600	500	600	500	1100	1600	1700	400	1200	0	1200	2400	700	400
21	100	0	0	200	100	100	200	400	300	1200	400	300	600	400	800	600	600	100	400	1200	0	1800	700	500
22	400	100	100	400	200	200	500	500	700	2600	1100	700	1300	1200	2600	1200	1700	300	1200	2400	1800	0	2100	1100
23	300	0	100	500	100	100	200	300	500	1800	1300	700	800	1100	1000	500	600	100	300	700	700	2100	0	700
24	100	0	0	200	0	100	100	200	200	800	600	500	700	400	400	300	300	0	100	400	500	1100	700	0

(b) The symmetrical OD matrix associated to the Sioux-Falls graph

Figure 20 – Sioux-Falls graph and associated OD matrix

For this instance, Table 21, obtained from a private communication with M. Karbstein, provides the data used in Borndörfer and Karbstein [2012] and Karbstein [2013].

Orig	Dest	Distance	Time
1	2	0.01	6
1	3	0.006	4
2	1	0.01	6
2	6	0.008	5
3	1	0.006	4
3	4	0.006	4
3	12	0.006	4
4	3	0.006	4
4	5	0.003	2
4	11	0.01	6
5	4	0.003	2
5	6	0.006	4
5	9	0.008	5
6	2	0.008	5
6	5	0.006	4
6	8	0.003	2
7	8	0.005	3
7	18	0.003	2
8	6	0.003	2
8	7	0.005	3
8	9	0.016	10
8	16	0.008	5
9	5	0.008	5
9	8	0.016	10
9	10	0.005	3
10	9	0.005	3
10	11	0.008	5
10	15	0.01	6
10	16	0.006	4
10	17	0.013	8
11	4	0.01	6
11	10	0.008	5
11	12	0.01	6
11	14	0.006	4
12	3	0.006	4
12	11	0.01	6
12	13	0.005	3
13	12	0.005	3

Orig	Dest	Distance	Time
13	24	0.006	4
14	11	0.006	4
14	15	0.008	5
14	23	0.006	4
15	10	0.01	6
15	14	0.008	5
15	19	0.005	3
15	22	0.005	3
16	8	0.008	5
16	10	0.006	4
16	17	0.003	2
16	18	0.005	3
17	10	0.013	8
17	16	0.003	2
17	19	0.003	2
18	7	0.003	2
18	16	0.005	3
18	20	0.006	4
19	15	0.005	3
19	17	0.003	2
19	20	0.006	4
20	18	0.006	4
20	19	0.006	4
20	21	0.01	6
20	22	0.008	5
21	20	0.01	6
21	22	0.003	2
21	24	0.005	3
22	15	0.005	3
22	20	0.008	5
22	21	0.003	2
22	23	0.006	4
23	14	0.006	4
23	22	0.006	4
23	24	0.003	2
24	13	0.006	4
24	21	0.005	3
24	23	0.003	2

Figure 21 – Arcs characteristics for the Sioux Falls instance

D Evolution of the bus network of Poitiers

D.1 Pre-restructuring network

The pre-restructuring network is composed of 19 lines (20 when counting the east section of the line 1 as a line by itself). According to the topology based rules defined in

2.1, their topology is classified in Table 22

Line	Connexion		Diffusion	
	Major	Meshing	Major	Secondary
Line 1	-	-	o	-
Line 1B	-	-	o	-
Line 1E	-	-	o	-
Line 2	o	-	-	-
Line 3	-	o	o	-
Line 4	o	-	-	-
Line 5A	-	-	o	-
Line 5B	-	o	-	-
Line 5C	-	-	o	-
Line 6	-	o	-	-
Line 7	-	o	o	-
Line 8	-	o	o	-
Line 9	-	o	-	-
Line 10	-	o	-	-
Line 11	-	-	o	-
Line 12	-	o	-	-
Line 13	-	-	o	-
Line 14	-	o	-	-
Citadine	-	o	-	-

Table 22 – Typology classification of the bus lines of the pre-restructuring network

D.2 Post-restructuring network

The post-restructuring network is composed of 22 lines (23 when counting the east section of the line 1 as a line by itself). Table 23 provides a breakdown of these lines according to whether they were already part of the pre-restructuring network or not. This allows us to see the extent to which the new network builds on the existing one or introduces entirely new lines. Moreover, according to topology based rules defined in 2.1, these bus lines are also classified in Table 23.

Line	Existing before restructuring	Name before restructuring	Connexion		Diffusion	
			Major	Meshing	Major	Secondary
Line 1	o	Line 1	-	-	o	-
Line 1E	o	Line 1E	-	-	o	-
Line 2	o	Line 2	o	-	-	-
Line 3	o	Line 4	o	-	-	-
Line 10	o	Line 10 (North)	-	o	-	-
Line 11	-	-	-	-	o	-
Line 12	-	-	-	o	-	-
Line 13	-	-	-	o	-	-
Line 14	o	Line 14	-	o	-	-
Line 15	-	-	-	o	-	-
Line 16	o	Line 8 (South)	-	o	-	-
Line 17	o	Line 7 (East)	-	o	-	-
Line 20	o	Line 13	-	-	o	-
Line 21	o	Line 1B	-	-	o	-
Line 22	-	-	-	-	o	-
Line 23	o	La Citadine	-	o	-	-
Line 24	o	Line 5C	-	-	o	-
Line 25	o	Bus 5B	-	o	-	-
Line 27	o	Route 10 (South)	-	o	-	-
Line 28	o	Line 5A	-	-	o	-
Line 29	o	Line 7 (West)	-	-	o	-
Line O	-	-	o	-	-	-

Table 23 – Typology classification of the bus lines of the post-restructuring network

E Proposed network

We propose a network, called Network *B*, that consists of 12 existing lines in the post-restructuring network, as well as 11 new lines. Among these new lines, 5 are designed to serve as major connections, 3 as meshing connections, and 3 as major diffusion lines. Figure 22 gives an overview of the proposed network *B*. The dashed lines represent the newly generated lines, while the solid lines represent the original post-restructuring network lines.

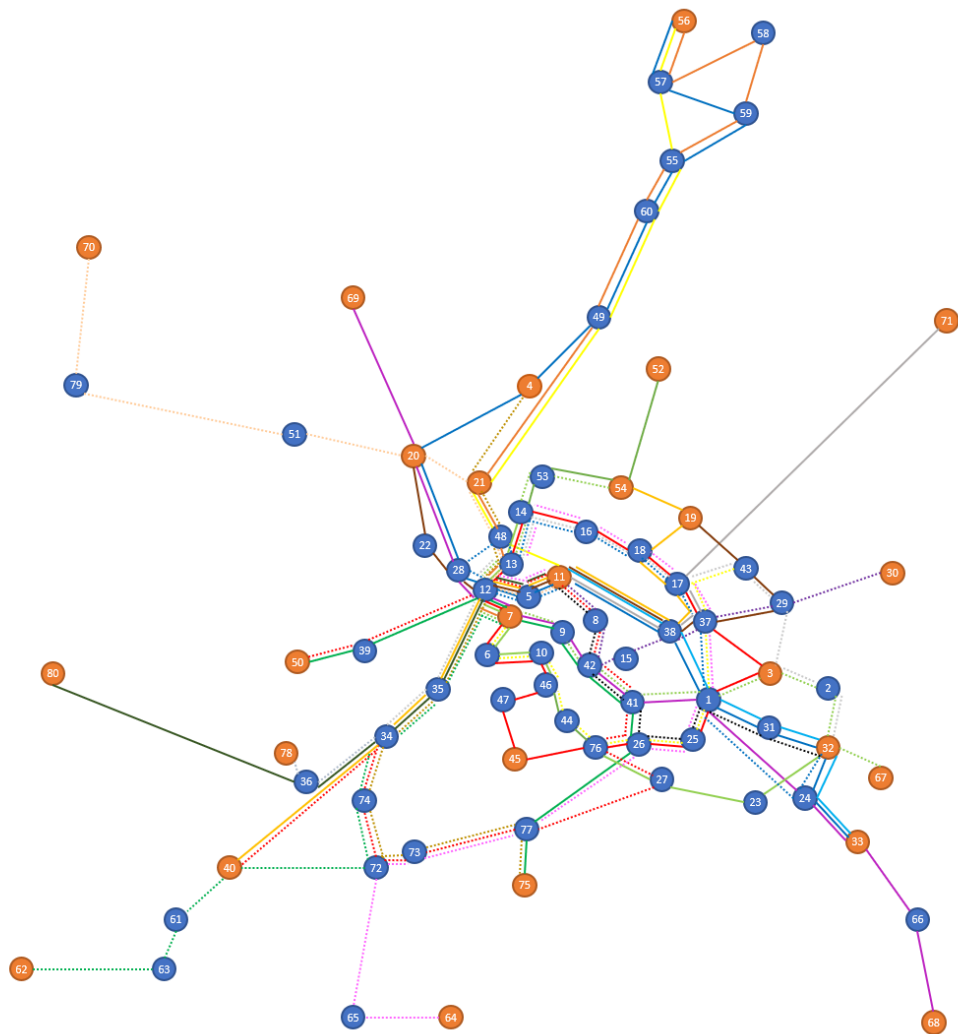


Figure 22 – Overview of the proposed network B

E.1 Post-restructuring network retained lines

Line index	Trip
Line 1	[33,24,32,31,1,37,38,11,5,12,28,20,4,49,60,55,59,57,56]
Line 1 east	[33,24,32,31,1,37,38,11,5]
Line 1e	[11, 5, 12, 13, 48, 21, 49, 60, 55, 57, 56]
Line 2	[3, 1, 25, 26, 76, 45, 47, 46, 10, 6, 7, 12, 13, 14, 16, 18, 17, 37, 3]
Line 3	[54, 19, 18, 17, 37, 38, 11, 5, 12, 35, 34, 40]
Line 11	[69, 20, 28, 12, 7, 9, 42, 41, 1, 24, 33, 66, 68]
Line 12	[50, 39, 12, 7, 9, 42, 41, 26, 77, 75]
Line 13	[52, 54, 53, 14, 13, 12, 7, 6, 10, 46, 44, 76, 27, 23, 32]
Line 16	[19, 43, 29, 37, 38, 11, 5, 12, 28, 22, 20]
Line 21	[11,5,12,13,48,21,49,60,55,59,58,57,56]
Line 22	[71,17,37,38,11]
Line 29	[11,5,12,35,34,36,80]

Table 24 – Post-restructuring network retained lines

E.2 Created Major connection lines

Line index	Trip
Line <i>new</i> ₁	[11, 8, 42, 41, 26, 25, 1, 31, 32]
Line <i>new</i> ₂	[32, 24, 1, 37, 17, 18, 16, 14, 13, 48, 28, 12, 5, 11]
Line <i>new</i> ₃	[30, 29, 37, 38, 15, 42, 8, 11, 5, 12, 7, 6, 10, 46, 47, 45]
Line <i>new</i> ₄	[19, 43, 17, 37, 1, 25, 26, 76, 44, 46, 10, 6, 7]
Line <i>new</i> ₅	[67, 32, 2, 3, 1, 41, 42, 9, 7, 12, 13, 14, 53, 54]

Table 25 – Major connexion lines created

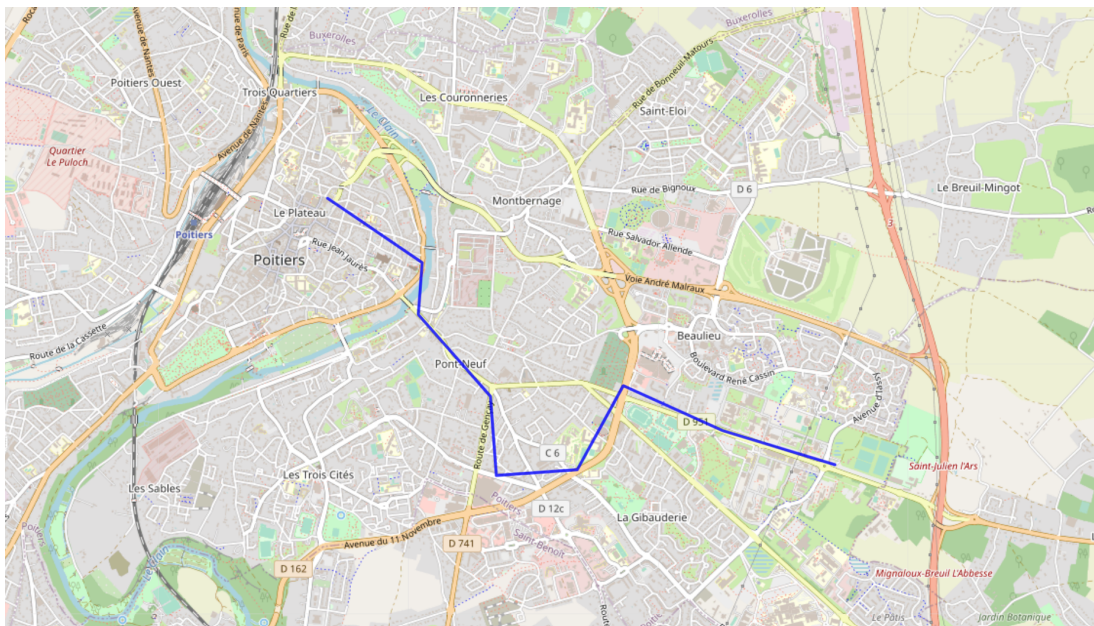


Figure 23 – Line new_1

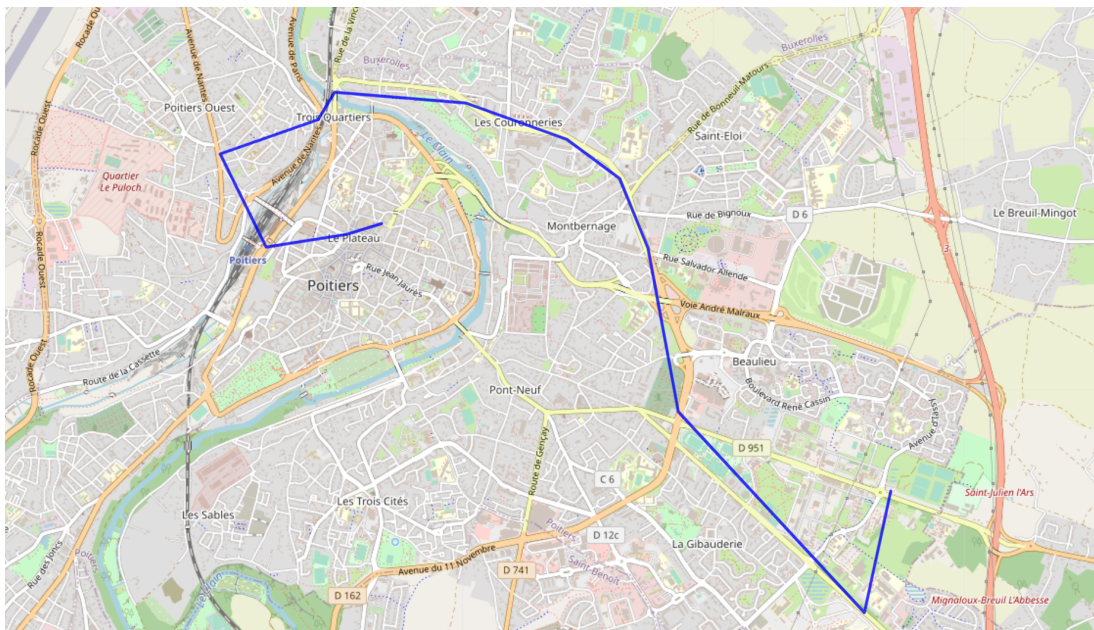


Figure 24 – Line new_2

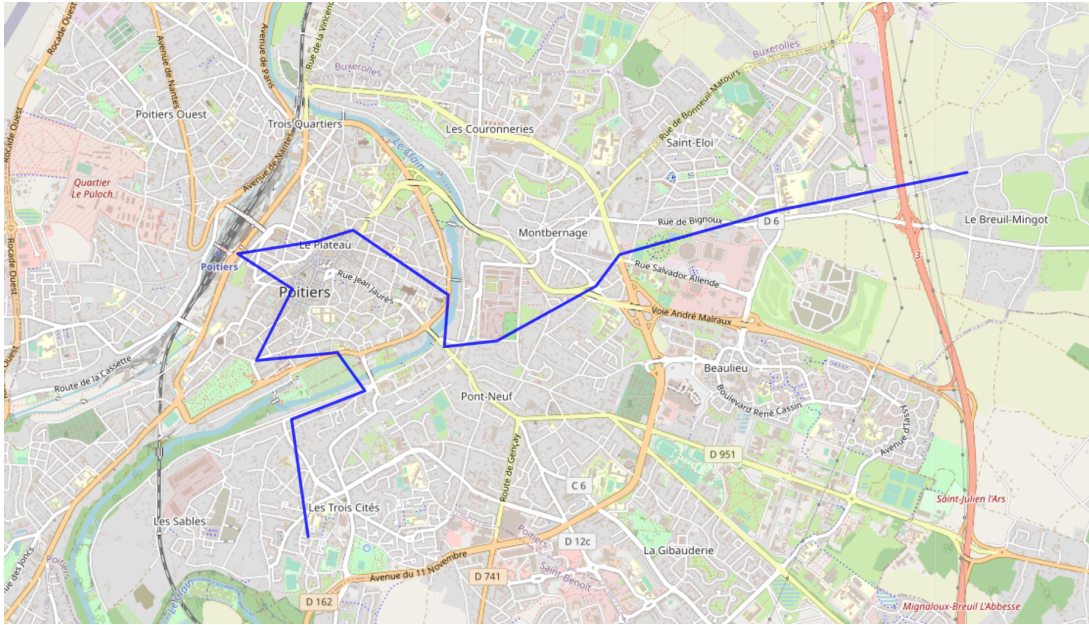


Figure 25 – Line *new3*

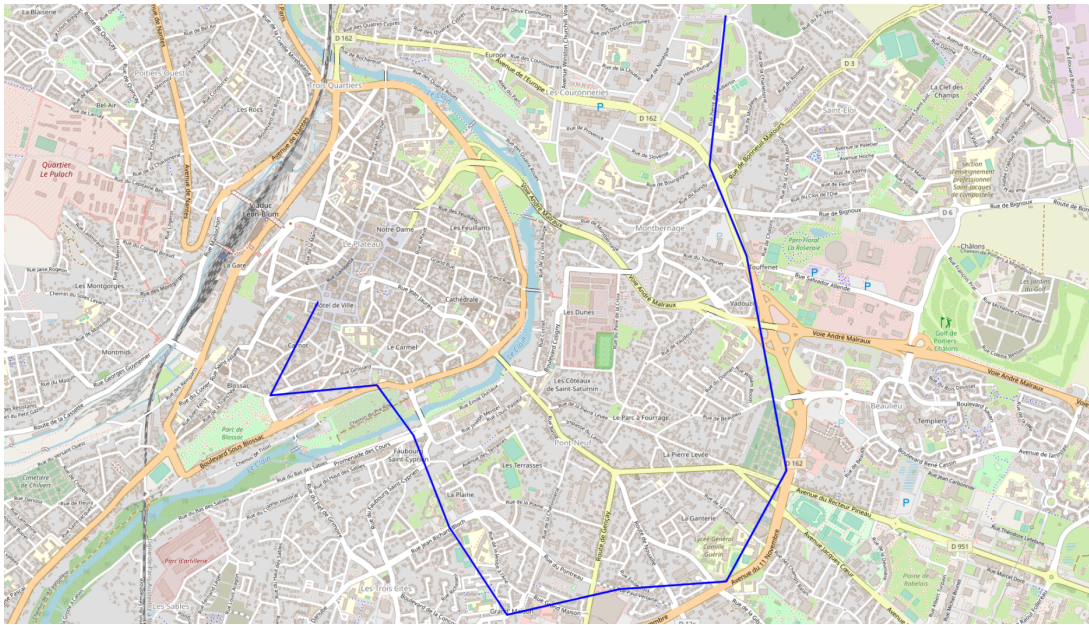


Figure 26 – Line *new4*

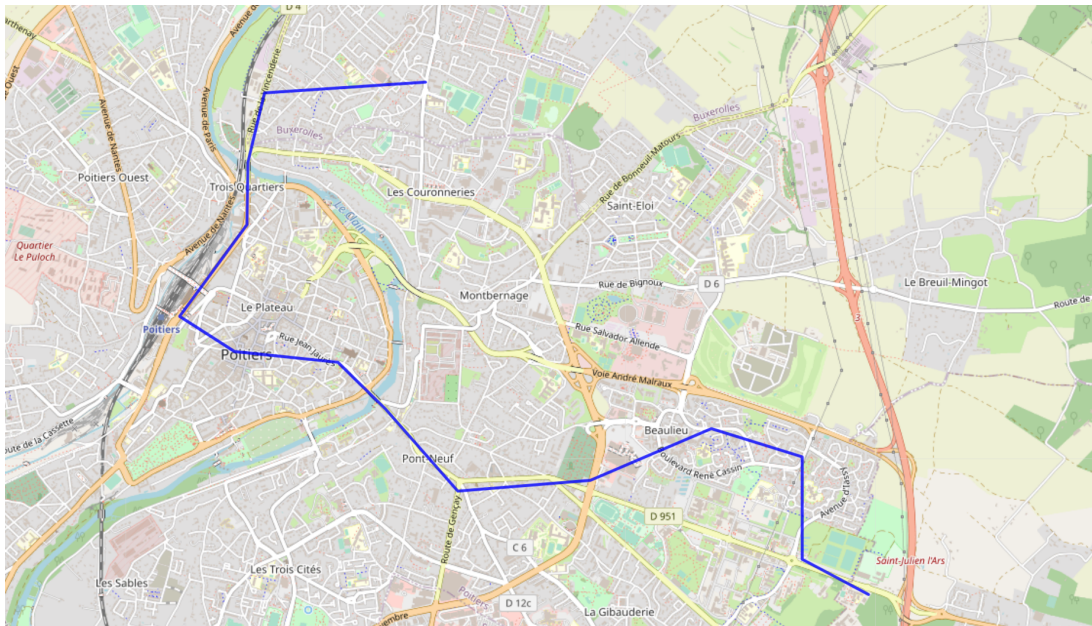


Figure 27 – Line new_5

E.3 Meshing connexion lines created

Line index	Trip
Line new_6	[40, 34, 74, 72, 73, 77, 27, 76, 26, 41, 42, 8, 11, 5, 12, 39, 50]
Line new_7	[78, 36, 34, 35, 12, 13, 14, 16, 18, 17, 43, 29, 3, 2, 32]
Line new_8	[75, 77, 73, 72, 74, 34, 35, 12, 13, 48, 21, 4]

Table 26 – Meshing connection lines created

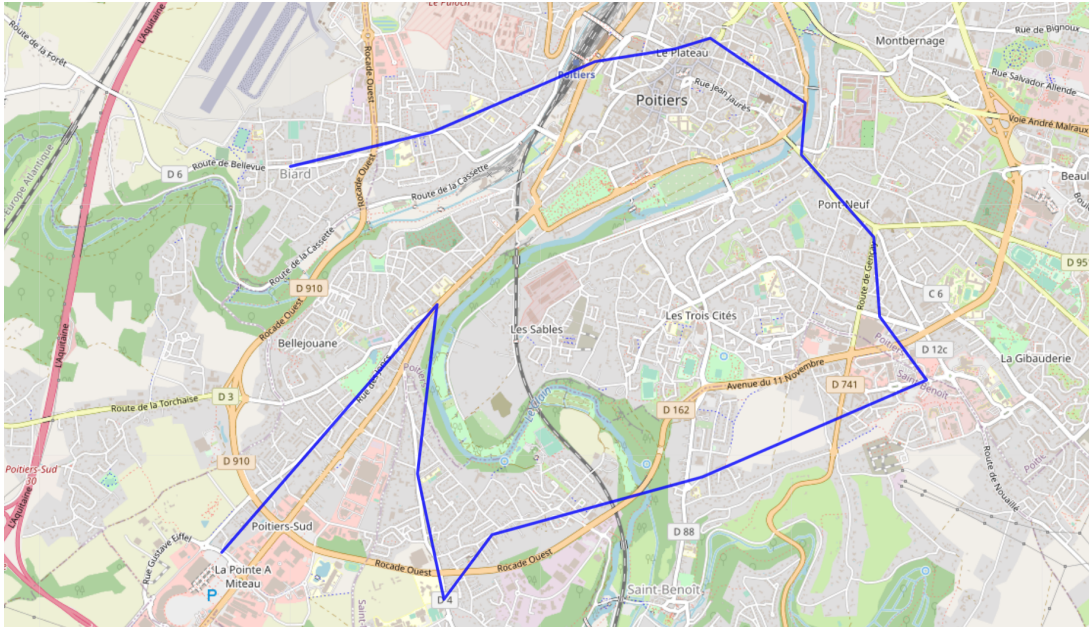


Figure 28 – Line *new*₆



Figure 29 – Line *new*₇

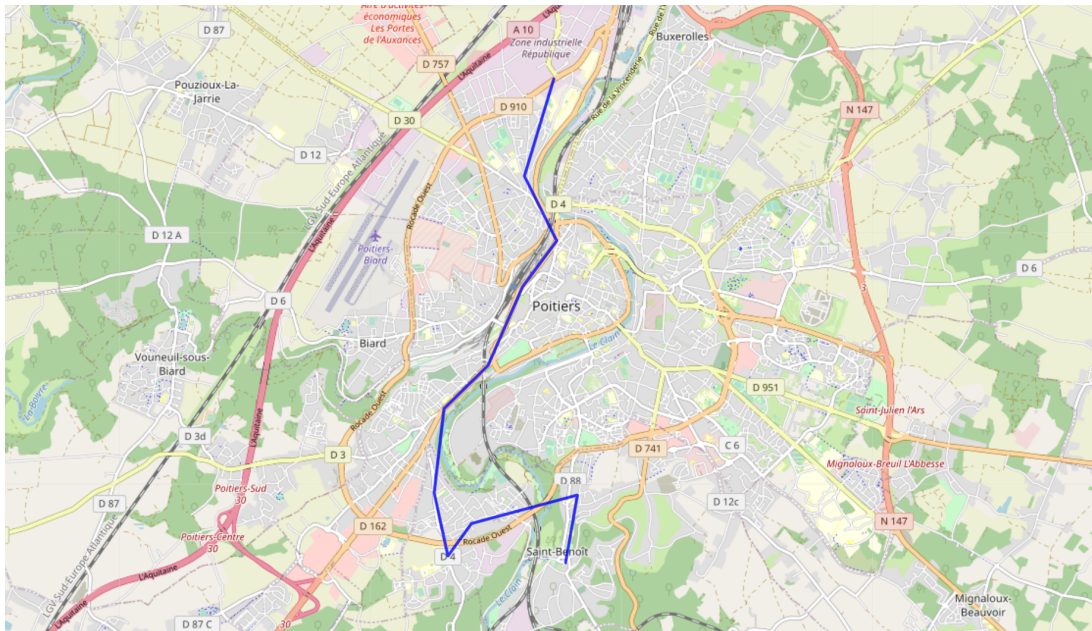


Figure 30 – Line *new*₈

E.4 Major diffusion lines created

Line index	Trip
Line <i>new</i> ₉	[7, 12, 35, 34, 74, 72, 40, 61, 63, 62]
Line <i>new</i> ₁₀	[11, 5, 12, 13, 14, 16, 18, 17, 37, 1, 25, 26, 77, 73, 72, 65, 64]
Line <i>new</i> ₁₁	[70, 79, 51, 20, 21, 48, 13, 12, 5, 11]

Table 27 – Major diffusion lines created

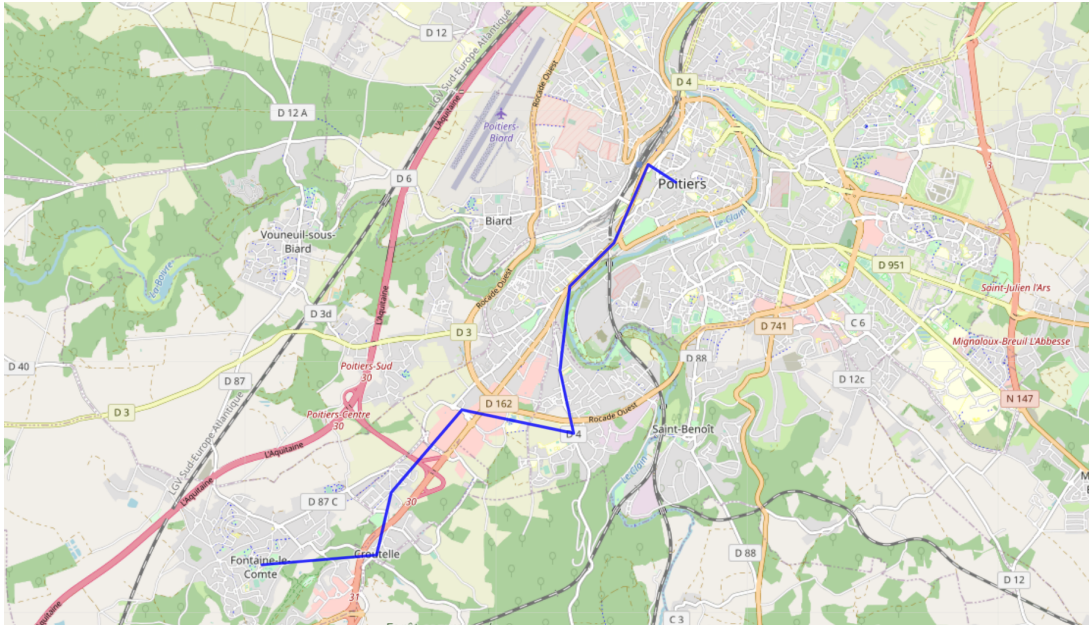


Figure 31 – Line *new*₉

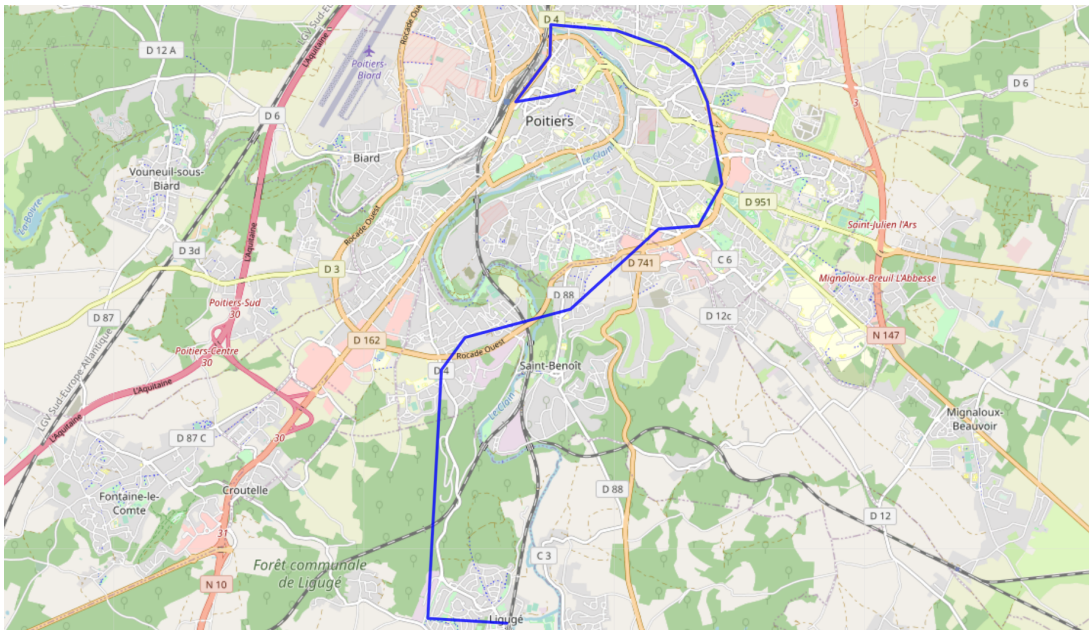


Figure 32 – Line *new*₁₀

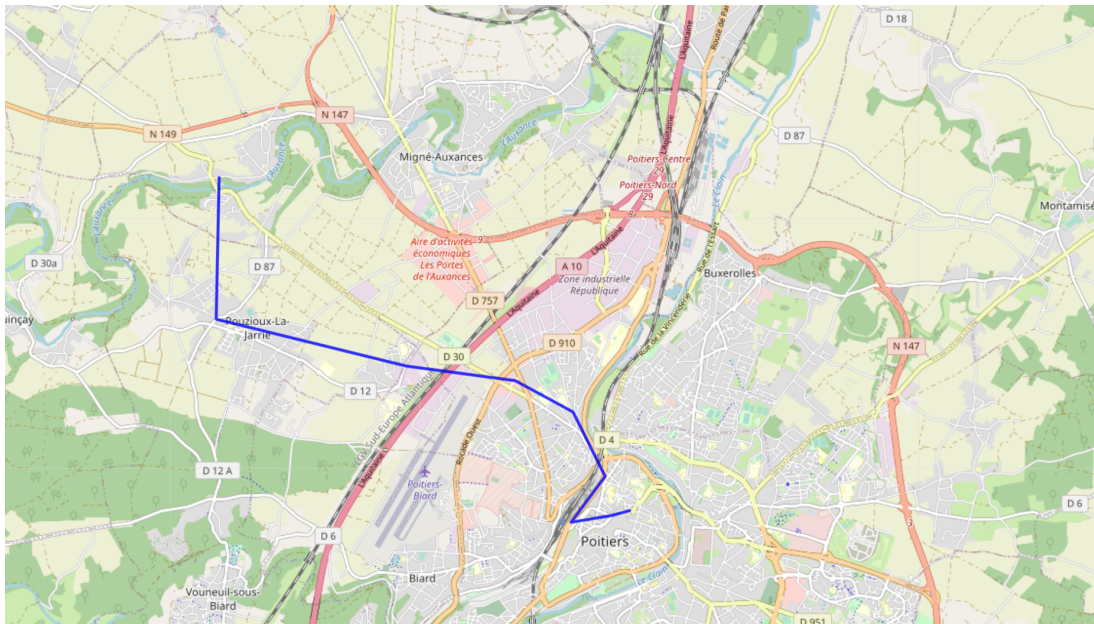


Figure 33 – Line *new11*

F Line frequencies

This section outlines the Monday-Friday line frequencies for the pre-restructuring, post-restructuring, and proposed network *B*. The two tables described in Subsections F.1 and F.2 list the line frequencies operated for the pre-restructuring network in the 2013-2014 and 2020-2021 seasons. Meanwhile, the two tables described in Subsections F.3 and F.4 present the line frequencies obtained with PPSP on the post-restructuring network and CGEP-proposed network *B*, respectively. For these two last tables, the line frequencies are based on the third passenger choice mode described in Chapter 5, which considers "passenger directness" during off-peak hours and "passenger shortest-path" during peak hours. Off-peak hours are defined as 8:00-12:00, 14:00-16:30, and 18:30-21:00, while peak hours are defined by the operator as 6:00-8:00, 12:00-14:00, and 16:30-18:30.

F.1 Operated line frequencies for the pre-restructuring RTP network

	6h-8h	8h-12h	12h- 14h	14h- 16h30	16h30- 18h30	18h30- 21h	Total day
Line 1	4 \ 3	12 \ 13	5 \ 7	8 \ 7	6 \ 6	5 \ 6	40 \ 42
Line 1 <i>b</i>	4 \ 4	2 \ 3	2 \ 2	2 \ 1	4 \ 2	3 \ 2	17 \ 14
Line 1 (east)	3 \ 7	9 \ 6	8 \ 7	2 \ 5	8 \ 7	6 \ 3	36 \ 35
Line 1 <i>e</i>	3 \ 0	3 \ 0	1 \ 0	0 \ 2	0 \ 3	0 \ 3	7 \ 8
Line 2	8 \ 4	25 \ 12	11 \ 7	14 \ 11	12 \ 9	15 \ 5	85 \ 48
Line 3	7 \ 6	11 \ 10	9 \ 8	6 \ 8	7 \ 8	6 \ 4	46 \ 44
Line 4	6 \ 4	18 \ 20	11 \ 11	10 \ 10	9 \ 8	7 \ 7	61 \ 60
Line 5 <i>a</i>	1 \ 3	2 \ 5	3 \ 3	4 \ 2	3 \ 2	4 \ 0	17 \ 15
Line 5 <i>b</i>	2 \ 3	6 \ 5	3 \ 3	3 \ 4	6 \ 4	4 \ 4	24 \ 23
Line 5 <i>c</i>	2 \ 2	1 \ 2	1 \ 2	1 \ 0	1 \ 2	2 \ 1	8 \ 9
Line 6	1 \ 2	3 \ 3	2 \ 3	3 \ 1	4 \ 5	2 \ 1	15 \ 15
Line 7	3 \ 5	9 \ 9	7 \ 6	5 \ 7	6 \ 6	6 \ 5	36 \ 38
Line 8	6 \ 5	14 \ 14	7 \ 7	8 \ 6	9 \ 9	6 \ 8	50 \ 49
Line 9	8 \ 4	16 \ 16	8 \ 8	9 \ 10	8 \ 7	5 \ 7	54 \ 52
Line 10	3 \ 2	5 \ 5	3 \ 3	2 \ 3	4 \ 3	2 \ 3	19 \ 19
Line 10 (north)	1 \ 2	8 \ 7	3 \ 3	4 \ 5	3 \ 4	5 \ 3	24 \ 24
Line 11	4 \ 3	10 \ 10	6 \ 6	7 \ 6	5 \ 7	6 \ 5	38 \ 37
Line 12	2 \ 2	9 \ 11	5 \ 5	8 \ 7	4 \ 5	4 \ 2	32 \ 32
Line 13	2 \ 0	1 \ 0	1 \ 1	0 \ 0	0 \ 3	0 \ 2	4 \ 6
Line 14	1 \ 4	8 \ 8	3 \ 3	6 \ 5	4 \ 4	4 \ 3	26 \ 27
Line <i>cita</i>	2 \ 2	4 \ 5	2 \ 2	3 \ 3	4 \ 3	3 \ 5	18 \ 20

F.2 Operated line frequencies for the post-restructuring RTP network

	6h-8h	8h-12h	12h-14h	14h-16h30	16h30-18h30	18h30-21h	Total day
Line 1	6 \ 4	9 \ 9	6 \ 6	5 \ 5	5 \ 5	7 \ 7	38 \ 36
Line 1 (east)	0 \ 3	9 \ 10	6 \ 5	6 \ 6	4 \ 4	1 \ 1	26 \ 29
Line 1e	2 \ 0	5 \ 0	1 \ 1	0 \ 1	0 \ 4	0 \ 2	8 \ 8
Line 2	6 \ 6	19 \ 19	9 \ 9	11 \ 11	9 \ 9	9 \ 9	63 \ 63
Line 3	5 \ 5	14 \ 14	5 \ 5	7 \ 6	5 \ 6	8 \ 7	44 \ 43
Line 3 (south)	0 \ 0	0 \ 0	3 \ 5	0 \ 0	4 \ 3	1 \ 2	8 \ 10
Line 10	4 \ 3	5 \ 5	3 \ 3	3 \ 3	4 \ 4	2 \ 2	21 \ 20
Line 11	5 \ 4	9 \ 9	6 \ 6	6 \ 5	6 \ 7	5 \ 5	37 \ 36
Line 12	1 \ 2	4 \ 7	3 \ 2	4 \ 4	4 \ 3	3 \ 3	19 \ 21
Line 12 (west)	1 \ 0	1 \ 0	0 \ 0	0 \ 0	0 \ 0	1 \ 1	3 \ 1
Line 12 (east)	1 \ 0	2 \ 0	1 \ 2	0 \ 0	1 \ 2	0 \ 1	5 \ 5
Line 13	5 \ 3	9 \ 9	6 \ 6	6 \ 5	5 \ 5	5 \ 5	36 \ 33
Line 14	1 \ 2	3 \ 4	2 \ 3	2 \ 1	3 \ 3	1 \ 1	12 \ 14
Line 15	4 \ 6	9 \ 9	6 \ 5	4 \ 5	7 \ 5	5 \ 5	35 \ 35
Line 16	5 \ 4	9 \ 10	6 \ 5	5 \ 6	5 \ 5	5 \ 4	35 \ 34
Line 17	4 \ 5	9 \ 7	5 \ 5	4 \ 3	4 \ 5	5 \ 4	31 \ 29
Line 20	2 \ 0	1 \ 0	1 \ 1	0 \ 0	0 \ 2	0 \ 2	4 \ 5
Line 21	1 \ 3	4 \ 4	2 \ 2	1 \ 1	4 \ 2	2 \ 2	14 \ 14
Line 22	3 \ 0	1 \ 0	1 \ 1	0 \ 0	2 \ 2	0 \ 2	7 \ 5
Line 23	1 \ 1	3 \ 0	1 \ 1	3 \ 0	1 \ 3	0 \ 1	9 \ 6
Line 24	1 \ 2	1 \ 2	1 \ 1	1 \ 0	1 \ 2	2 \ 0	7 \ 7
Line 25	0 \ 1	3 \ 5	1 \ 2	2 \ 3	3 \ 4	2 \ 0	11 \ 15
Line 27	2 \ 2	2 \ 4	1 \ 1	1 \ 1	3 \ 2	2 \ 1	11 \ 11
Line 28	0 \ 3	0 \ 1	1 \ 1	0 \ 0	2 \ 1	2 \ 0	5 \ 6
Line 29	0 \ 2	0 \ 2	1 \ 1	2 \ 0	1 \ 2	2 \ 0	6 \ 7
Line <i>O</i>	2 \ 0	8 \ 0	4 \ 0	5 \ 0	4 \ 0	5 \ 0	28 \ 0

F.3 PPSP-proposed line frequencies for the post-restructuring RTP network

	6h-8h	8h-12h	12h-14h	14h-16h30	16h30-18h30	18h30-21h	Total day
Line 1	3 \ 2	3 \ 5	2 \ 4	2 \ 4	5 \ 7	1 \ 3	16 \ 25
Line 1 (east)	10 \ 8	13 \ 11	8 \ 6	6 \ 4	6 \ 4	4 \ 2	47 \ 35
Line 1e	4 \ 6	6 \ 8	1 \ 2	3 \ 2	5 \ 3	1 \ 1	20 \ 22
Line 2	10 \ 8	12 \ 10	9 \ 7	9 \ 7	8 \ 6	2 \ 3	50 \ 41
Line 3	2 \ 4	6 \ 8	3 \ 4	4 \ 6	4 \ 3	2 \ 2	21 \ 27
Line 10	2 \ 2	3 \ 5	3 \ 2	5 \ 3	3 \ 2	1 \ 1	17 \ 15
Line 11	7 \ 9	3 \ 5	2 \ 4	4 \ 2	4 \ 3	1 \ 1	21 \ 24
Line 12	2 \ 4	3 \ 5	1 \ 3	2 \ 4	3 \ 4	1 \ 2	12 \ 22
Line 13	4 \ 4	5 \ 6	5 \ 3	4 \ 5	4 \ 4	2 \ 2	24 \ 24
Line 14	2 \ 1	2 \ 1	1 \ 1	1 \ 1	1 \ 2	1 \ 1	8 \ 7
Line 15	2 \ 3	5 \ 4	4 \ 2	4 \ 2	6 \ 4	3 \ 1	24 \ 16
Line 16	4 \ 3	5 \ 4	4 \ 2	6 \ 4	6 \ 4	2 \ 2	27 \ 19
Line 17	3 \ 1	1 \ 2	1 \ 1	1 \ 1	1 \ 2	1 \ 1	8 \ 8
Line 20	1 \ 0	1 \ 0	1 \ 1	0 \ 0	0 \ 1	0 \ 1	3 \ 3
Line 21	2 \ 1	1 \ 1	1 \ 1	1 \ 2	2 \ 4	1 \ 1	8 \ 10
Line 22	1 \ 2	1 \ 3	1 \ 1	0 \ 0	3 \ 1	0 \ 0	6 \ 7
Line 23	1 \ 1	3 \ 2	1 \ 2	3 \ 2	2 \ 1	1 \ 1	11 \ 9
Line 24	1 \ 1	1 \ 0	1 \ 1	0 \ 0	1 \ 1	1 \ 0	5 \ 3
Line 25	3 \ 1	6 \ 4	2 \ 2	3 \ 4	1 \ 3	1 \ 1	16 \ 15
Line 27	3 \ 5	3 \ 3	2 \ 2	3 \ 2	2 \ 2	2 \ 1	15 \ 15
Line 28	2 \ 1	0 \ 1	1 \ 1	0 \ 1	1 \ 3	1 \ 0	5 \ 7
Line 29	0 \ 0	0 \ 1	1 \ 1	1 \ 0	1 \ 3	1 \ 0	4 \ 5
Line O	2 \ 0	5 \ 0	3 \ 0	4 \ 0	4 \ 0	2 \ 0	20 \ 0

F.4 PPSP-proposed line frequencies for the CGEP-proposed network B

	6h-8h	8h-12h	12h-14h	14h-16h30	16h30-18h30	18h30-21h	Total day
Line 1	4 \ 2	4 \ 3	3 \ 4	2 \ 3	4 \ 6	1 \ 2	18 \ 20
Line 1 (east)	5 \ 3	7 \ 7	4 \ 4	3 \ 4	3 \ 4	2 \ 3	24 \ 25
Line 1e	5 \ 7	6 \ 8	2 \ 2	4 \ 2	8 \ 6	2 \ 1	27 \ 26
Line 2	3 \ 2	4 \ 6	3 \ 3	3 \ 5	4 \ 6	1 \ 3	18 \ 25
Line 3	2 \ 4	6 \ 8	3 \ 4	7 \ 6	5 \ 4	2 \ 1	25 \ 27
Line 11	3 \ 5	3 \ 4	1 \ 2	4 \ 2	4 \ 2	1 \ 1	16 \ 16
Line 12	1 \ 3	2 \ 2	1 \ 1	2 \ 2	3 \ 1	1 \ 1	10 \ 10
Line 13	3 \ 5	2 \ 2	2 \ 2	3 \ 2	3 \ 2	1 \ 1	14 \ 14
Line 16	2 \ 3	4 \ 4	3 \ 2	6 \ 4	5 \ 4	2 \ 2	22 \ 19
Line 21	2 \ 1	2 \ 2	1 \ 1	2 \ 1	2 \ 2	1 \ 1	10 \ 8
Line 22	1 \ 2	1 \ 3	1 \ 1	0 \ 0	3 \ 1	0 \ 1	6 \ 8
Line 29	3 \ 1	1 \ 1	1 \ 1	1 \ 1	1 \ 3	1 \ 1	8 \ 8
Line new_1	1 \ 2	1 \ 2	1 \ 2	1 \ 1	1 \ 2	1 \ 1	6 \ 10
Line new_2	4 \ 2	4 \ 2	3 \ 2	3 \ 3	4 \ 3	1 \ 1	19 \ 13
Line new_3	3 \ 2	5 \ 3	2 \ 2	4 \ 3	3 \ 2	1 \ 1	18 \ 13
Line new_4	3 \ 3	4 \ 5	2 \ 3	4 \ 3	2 \ 4	1 \ 2	16 \ 20
Line new_5	3 \ 1	4 \ 4	1 \ 3	2 \ 3	2 \ 3	1 \ 2	13 \ 16
Line new_6	2 \ 2	2 \ 3	1 \ 2	2 \ 2	2 \ 2	1 \ 1	10 \ 12
Line new_7	1 \ 2	3 \ 3	2 \ 2	3 \ 2	4 \ 2	1 \ 1	14 \ 12
Line new_8	1 \ 1	1 \ 1	1 \ 1	1 \ 1	1 \ 1	1 \ 1	6 \ 6
Line new_9	5 \ 3	6 \ 4	3 \ 2	3 \ 1	3 \ 3	1 \ 0	21 \ 13
Line new_{10}	1 \ 1	1 \ 1	1 \ 1	1 \ 1	1 \ 1	1 \ 1	6 \ 6
Line new_{11}	1 \ 1	2 \ 2	1 \ 1	2 \ 1	1 \ 2	1 \ 1	8 \ 8

BIBLIOGRAPHY

- En-13816. URL <https://www.en-standard.eu/din-en-13816>.
- United nations, department of economic and social affairs, population division (2018). the world's cities in 2018. 2018.
- Anne Aguilera-Belanger, Gilles Cabadi, Benoit Conti, and Zembri-Mary Genevieve. *Concevoir Une Offre de transport public*. 2017.
- Renato Oliveira Arbex and Claudio Barbieri da Cunha. Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, 81:355–376, November 2015. doi: 10.1016/j.trb.2015.06.014.
- Agustin Arizti, Antonio Mauttone, and Maria E. Urquhart. A Bilevel Approach to Frequency Optimization in Public Transportation Systems. 2018. doi: 10.4230/OASICS.ATMOS.2018.7.
- Mohamed Hadi Baaj and Hani S. Mahmassani. TRUST: A LISP Program for the Analysis of Transit Route Configurations. *Transportation Research Record*, page 11, 1990.
- Mohamed Hadi Baaj and Hani S. Mahmassani. Hybrid route generation heuristic algorithm for the design of transit networks. *Transportation Research Part C: Emerging Technologies*, 3:31–50, February 1995. doi: 10.1016/0968-090X(94)00011-S.
- Roberto Baldacci, Nicos Christofides, and Aristide Mingozzi. An exact algorithm for the vehicle routing problem based on the set partitioning formulation with additional cuts. *Mathematical Programming*, 115:351–385, 2008. doi: 10.1007/s10107-007-0178-5.
- Hillel Bar-Gera. Transportation network test problems, 2021. URL <https://github.com/bstabler/TransportationNetworks>.
- Dimitris Bertsimas, Yee Sian Ng, and Julia Yan. Data-driven transit network design at scale. *Operations Research*, 69(4):1118–1133, 2021. doi: 10.1287/opre.2020.2057.

- Ralf Borndörfer and Marika Karbstein. A direct connection approach to integrated line planning and passenger routing. In *12th Workshop on algorithmic approaches for transportation modelling, optimization, and systems*, 2012. doi: 10.4230/OASICS.ATMOS.2012.47.
- Ralf Borndörfer, Martin Grötschel, and Marc E. Pfetsch. A column-generation approach to line planning in public transport. *Transportation Science*, 41:123–132, 2007. doi: 10.1287/trsc.1060.0161.
- Ralf Borndörfer, Marika Neumann, and Marc E Pfetsch. The line connectivity problem. In *Operations Research Proceedings 2008*, pages 557–562. Springer, 2009. doi: 10.1007/978-3-642-00142-0_90.
- Albert Bouma and Christine Oltrogge. Linienplanung und simulation für öffentliche verkehrswege in praxis und theorie. *ETR. Eisenbahntechnische Rundschau*, 43:369–378, 1994.
- Lucile Brethomé, Rémy Chevrier, Niels van Oort, and Joaquin Rodriguez. Passenger-oriented optimization of lines in a mass transit system. *7th International Seminar on Railway Operations Modelling and Analysis*, page 9, 2018.
- Michael R. Bussieck, Peter Kreuzer, and Uwe T. Zimmermann. Optimal lines for railway systems. *European Journal of Operational Research*, 96:54–63, 1997. doi: 10.1016/0377-2217(95)00367-3.
- Rosalia Camporeale, Leonardo Caggiani, Achille. Fonzone, and Michele Ottomanelli. Quantifying the impacts of horizontal and vertical equity in transit route planning. *Transportation Planning and Technology*, 40:28–44, January 2017. doi: 10.1080/03081060.2016.1238569.
- Rosalia Camporeale, Leonardo Caggiani, and Michele Ottomanelli. Modeling horizontal and vertical equity in the public transport design problem: A case study. *Transportation Research Part A: Policy and Practice*, 125:184–206, July 2019. doi: 10.1016/j.tra.2018.04.006.
- Avishai Ceder and Nigel H. M. Wilson. Bus network design. *Transportation Research Part B: Methodological*, 20:331–344, 1986. doi: 10.1016/0191-2615(86)90047-0.

- Cerema. Transports collectifs urbains de province évolution 2011 - 2016 - annuaire statistique. 2019. URL <https://www.cerema.fr/fr/centre-ressources/boutique/transports-collectifs-urbains-province-evolution-2011-2016>.
- Ernesto Cipriani, Marco Petrelli, and Gaetano Fusco. A multimodal transit network design procedure for urban areas. *Advances in Transportation Studies*, pages 5–20, 01 2006.
- Ernesto Cipriani, Stefano Gori, and Marco Petrelli. Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies*, 20:3–14, February 2012. doi: 10.1016/j.trc.2010.09.003.
- Isabelle Constantin and Michael Florian. Optimizing frequencies in a transit network: a nonlinear bi-level programming approach. *International Transactions in Operational Research*, 2:149–164, 1995. doi: 10.1016/0969-6016(94)00023-M.
- Joaquin De Cea and Enrique Fernandez. Transit Assignment for Congested Public Transport Systems: An Equilibrium Model. *Transportation Science*, 27:133–147, 1993. doi: 10.1287/trsc.27.2.133.
- Stephan Dempe and Floriane Mefo Kue. Solving discrete linear bilevel optimization problems using the optimal value reformulation. *Journal of Global Optimization*, 68: 255–277, June 2017. doi: 10.1007/s10898-016-0478-5.
- Guy Desaulniers, Jacques Desrosiers, and Marius M Solomon. *Column generation*, volume 5. 2006.
- Javier Duran-Micco, Evert Vermeir, and Pieter Vansteenwegen. Considering emissions in the transit network design and frequency setting problem with a heterogeneous fleet. *European Journal of Operational Research*, 282:580–592, April 2020. doi: 10.1016/j.ejor.2019.09.050.
- Javier Durán-Micco and Pieter Vansteenwegen. A survey on the transit network design and frequency setting problem. *Public Transport*, October 2021. doi: 10.1007/s12469-021-00284-y.
- Javier Durán-Micco, Marcel van Kooten Niekerk, and Pieter Vansteenwegen. Designing bus line plans for realistic cases - the Utrecht case study. *Expert Systems with Applications*, 187, January 2022. doi: 10.1016/j.eswa.2021.115918.

- Bruno Faivre d'Arcier. Measuring the performance of urban public transport in relation to public policy objectives. *Research in Transportation Economics*, 48:67–76, December 2014. doi: 10.1016/j.retrec.2014.09.033.
- Dominique Feillet. A tutorial on column generation and branch-and-price for vehicle routing problems. *4OR*, 8:407–424, December 2010. doi: 10.1007/s10288-010-0130-z.
- Pierre Fouilhoux, Omar Jorge Ibarra-Rojas, Safia Kedad-Sidhoum, and Yasmin A Rios-Solis. Valid inequalities for the synchronization bus timetabling problem. *European Journal of Operational Research*, 251:442–450, 2016. doi: 10.1016/j.ejor.2015.12.006.
- Hector Gatt, Jean-Marie Freche, Fabien Lehuédé, and Thomas G Yeung. A Column Generation-Based Heuristic for the Line Planning Problem with Service Levels. In *ATMOS 2021: International Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, Lisbon, Portugal, September 2021. doi: 10.4230/OASICS.ATMOS.2021.19.
- Hector Gatt, Jean-Marie Freche, Fabien Lehuédé, and Arnaud Laurent. A Bilevel Model for the Frequency Setting Problem. In *ATMOS 2022 : International Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, Potsdam, Germany, September 2022a. doi: 10.4230/OASICS.ATMOS.2022.5.
- Hector Gatt, Jean-Marie Freche, Fabien Lehuédé, and Thomas G. Yeung. Génération de Colonnes pour le Problème de Conception de Lignes de Bus avec Niveaux de Service. In *ROADEF 2022 : 23ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision*, Villeurbanne - Lyon, France, February 2022b. INSA Lyon.
- Hector Gatt, Jean-Marie Freche, Arnaud Laurent, and Fabien Lehuédé. Optimisation de la fréquences des lignes sur un réseau de bus. In *ROADEF 2023 : 24ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision*, Rennes, France, February 2023. Rennes Business School.
- Philine Gattermann, Jonas Harbering, and Anita Schöbel. Line pool generation. *Public Transport*, 9:7–32, 2017. doi: 10.1007/s12469-016-0127-x.
- Ricardo Giesen, Héctor Martínez, Antonio Mauttone, and María E. Urquhart. A method for solving the multi-objective transit frequency optimization problem. *Journal of Advanced Transportation*, 50:2323–2337, December 2016. doi: 10.1002/atr.1461.

- Estèle Glize, Roberto Roberti, Nicolas Jozefowicz, and Sandra Ulrich Ngueveu. Exact methods for mono-objective and Bi-Objective Multi-Vehicle Covering Tour Problems. *European Journal of Operational Research*, 283:812–824, June 2020. doi: 10.1016/j.ejor.2019.11.045.
- Marc Goerigk and Marie Schmidt. Line planning with user-optimal route choice. *European Journal of Operational Research*, 259:424–436, June 2017. doi: 10.1016/j.ejor.2016.10.034.
- Valérie Guihaire and Jin-Kao Hao. *Transit network design and scheduling: A global review*. PhD thesis, 2008.
- Gabriel Gutiérrez-Jarpa, Gilbert Laporte, Vladimir Marianov, and Luigi Moccia. Multi-objective rapid transit network design with modal competition: The case of Concepción, Chile. *Computers & Operations Research*, 78:27–43, February 2017. doi: 10.1016/j.cor.2016.08.009.
- Anthony F. Han and Nigel H. M. Wilson. The allocation of buses in heavily utilized networks with overlapping routes. *Transportation Research Part B: Methodological*, 16: 221–232, June 1982.
- Irene Heinrich, Philine Schiewe, and Constantin Seebach. Line Planning in Public Transport: Bypassing Line Pool Generation. May 2022. doi: 10.48550/arXiv.2205.09580.
- Omar Jorge Ibarra-Rojas, Felipe Delgado, Ricardo Giesen, and Juan Carlos Muñoz. Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B: Methodological*, 77:38–75, July 2015. doi: 10.1016/j.trb.2015.03.002.
- Shashi Bhushan Jha, Jitendra Kumar Jha, and Manoj Kumar Tiwari. A multi-objective meta-heuristic approach for transit network design and frequency setting problem in a bus transit system. *Computers & Industrial Engineering*, 130:166–186, April 2019. doi: 10.1016/j.cie.2019.02.025.
- Jian Gang Jin, Kwong Meng Teo, and Amedeo R. Odoni. Optimizing Bus Bridging Services in Response to Disruptions of Urban Transit Rail Networks. *Transportation Science*, 50:790–804, August 2016. doi: 10.1287/trsc.2014.0577.

- Marika Karbstein. *Line Planning and Connectivity*. PhD thesis, Mathematik und Naturwissenschaften der Technischen Universität Berlin, 2013.
- Tama Leventhal, George L. Nemhauser, and Leslie E. Trotter Jr. A column generation algorithm for optimal traffic assignment. *Transportation Science*, 7:168–176, 1973. doi: 10.1287/trsc.7.2.168.
- Jinpeng Liang, Jianjun Wu, Ziyou Gao, Huijun Sun, Xin Yang, and Hong K. Lo. Bus transit network design with uncertainties on the basis of a metro network: A two-step model framework. *Transportation Research Part B: Methodological*, 126:115–138, August 2019. doi: 10.1016/j.trb.2019.05.011.
- Ziyi Ma and Joseph Y. J. Chow. Transit Network Frequency Setting With Multi-Agent Simulation to Capture Activity-Based Mode Substitution. *Transportation Research Record: Journal of the Transportation Research Board*, November 2021. doi: 10.1177/036119812111056909.
- Christoph E. Mandl. Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research*, 5:396–404, 1980. doi: 10.1016/0377-2217(80)90126-5.
- Guillaume Marquez. *Problèmes de tournées de véhicules sur deux niveaux pour la logistique urbaine*. PhD thesis, Université de Bordeaux, 2020.
- Héctor Martínez, Antonio Mauttone, and María E. Urquhart. Frequency optimization in public transportation systems: Formulation and metaheuristic approach. *European Journal of Operational Research*, 236:27–36, 2014. doi: 10.1016/j.ejor.2013.11.007. Publisher: Elsevier.
- Ángel G. Marín. An extension to rapid transit network design problem. *TOP*, 15:231–241, December 2007. doi: 10.1007/s11750-007-0017-0.
- Ángel G. Marín and Patricia Jaramillo. Urban rapid transit network design: accelerated Benders decomposition. *Annals of Operations Research*, 169:35–53, July 2009. doi: 10.1007/s10479-008-0388-0.
- Antonio Mauttone and María E. Urquhart. A multi-objective metaheuristic approach for the Transit Network Design Problem. *Public Transport*, 1:253–273, November 2009. doi: 10.1007/s12469-010-0016-7.

- Christine L. Mumford. New heuristic and evolutionary operators for the multi-objective urban transit routing problem. In *2013 IEEE Congress on Evolutionary Computation*, pages 939–946. IEEE, June 2013. doi: 10.1109/CEC.2013.6557668.
- Gordon F. Newell. Dispatching Policies for a Transportation Route. *Transportation Science*, 5:91–105, February 1971. doi: 10.1287/trsc.5.1.91.
- Miloš Nikolić and Dušan Teodorović. A simultaneous transit network design and frequency setting: Computing with bees. *Expert Systems with Applications*, 41:7200–7209, November 2014. doi: 10.1016/j.eswa.2014.05.034.
- Yolanda Noriega and Michael Florian. L’Optimisation Des Fréquences D’un Réseau De Transport En Commun Moyennement Congestionné. *INFOR: Information Systems and Operational Research*, 41:129–153, May 2003. doi: 10.1080/03155986.2003.11732672.
- Emile Quinet. L’évaluation socioéconomique des investissements publics (tome1), 2013. URL <https://www.strategie.gouv.fr/publications/levaluation-socioeconomique-investissements-publics-tome1>.
- Hannah Ritchie and Max Roser. Urbanization. *Our World in Data*, 2018.
- Francisco Ruisanchez, Luigi dell’Olio, and Angel Ibeas. Design of a tabu search algorithm for assigning optimal bus sizes and frequencies in urban transport services: TABU SEARCH FOR ASSIGNING BUS SIZE AND FREQUENCY. *Journal of Advanced Transportation*, 46:366–377, October 2012. doi: 10.1002/atr.1195.
- Martin Schmidt and Yasmine Beck. A gentle and incomplete introduction to bilevel optimization, 2021. URL www.optimization-online.org/DB_FILE/2021/06/8450.pdf.
- Anita Schöbel. Line planning in public transportation: models and methods. *OR spectrum*, 34(3):491–510, 2012. doi: 10.1007/s00291-011-0251-6.
- Anita Schöbel and Susanne Scholl. Line planning with minimal traveling time. In *5th Workshop on Algorithmic Methods and Models for Optimization of Railways (ATMOS’05)*, 2006. doi: 10.4230/OASICS.ATMOS.2005.660.
- Prabhat Shrivastava and Margaret O’Mahony. A model for development of optimized feeder routes and coordinated schedules—A genetic algorithms approach. *Transport Policy*, 13:413–425, September 2006. doi: 10.1016/j.tranpol.2006.03.002.

- Heinz Spiess and Michael Florian. Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological*, 23:83–102, April 1989. doi: 10.1016/0191-2615(89)90034-9.
- Hemant K. Suman and Nomes B. Bolia. Improvement in direct bus services through route planning. *Transport Policy*, 81:263–274, September 2019. doi: 10.1016/j.tranpol.2019.07.001.
- Hemant Kumar Suman, Nomes B. Bolia, and Geetam Tiwari. Perception of potential bus users and impact of feasible interventions to improve quality of bus services in Delhi. *Case Studies on Transport Policy*, 6:591–602, December 2018. doi: 10.1016/j.cstp.2018.07.009.
- Weitiao Wu, Ronghui Liu, Wenzhou Jin, and Changxi Ma. Stochastic bus schedule coordination considering demand assignment and rerouting of passengers. *Transportation Research Part B: Methodological*, 121:275–303, March 2019. doi: 10.1016/j.trb.2019.01.010.
- Julia Yan. *From Data to Decisions in Urban Transit and Logistics*. PhD thesis, Massachusetts Institute of Technology, 2020.
- Zhongzhen Yang, Bin Yu, and Chuntian Cheng. A Parallel Ant Colony Algorithm for Bus Network Optimization. *Computer-Aided Civil and Infrastructure Engineering*, 22: 44–55, January 2007. doi: 10.1111/j.1467-8667.2006.00469.x.
- Bin Yu, Zhongzhen Yang, and Jinbao Yao. Genetic algorithm for bus frequency optimization. *Journal of Transportation Engineering*, 136:576–583, June 2010. doi: 10.1061/(ASCE)TE.1943-5436.0000119.
- Bin Yu, Zhong-Zhen Yang, Peng-Huan Jin, Shan-Hua Wu, and Bao-Zhen Yao. Transit route network design-maximizing direct and transfer demand density. *Transportation Research Part C: Emerging Technologies*, 22:58–75, June 2012. doi: 10.1016/j.trc.2011.12.003.
- Fang Zhao and Ike Ubaka. Transit Network Optimization - Minimizing Transfers and Optimizing Route Directness. *Journal of Public Transportation*, 7:63–82, March 2004. doi: 10.5038/2375-0901.7.1.4.

PERSONAL PUBLICATIONS LIST

Submitted to an international journal

- H. Gatt, J.M. Freche, F. Lehuédé, T. G. Yeung. Solving the Line Planning Problem with Service-Levels using a Column Generation-based Heuristic Algorithm.
- H. Gatt, J.M. Freche, A. Laurent, F. Lehuédé. A Matheuristic Approach to the Bi-level Frequency Setting Problem.

International conferences with proceedings

- H. Gatt, J.M. Freche, F. Lehuédé and T. G. Yeung. A Column Generation-Based Heuristic for the Line Planning Problem with Service Levels. ATMOS 2021: International Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. Lisbon, Portugal, 2021.
- H. Gatt, J.M. Freche, F. Lehuédé and A. Laurent. A Bilevel Model for the Frequency Setting Problem. ATMOS 2022 : International Symposium on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. Potsdam, Germany, 2022.

National Conferences

- H. Gatt, J.M. Freche, F. Lehuédé and T. G. Yeung. Génération de Colonnes pour le Problème de Conception de Lignes de Bus avec Niveaux de Service. ROADEF 2022 : 23 ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision. Villeurbanne - Lyon, France.
- H. Gatt, J.M. Freche, F. Lehuédé and A. Laurent. Optimisation de la Fréquence des Lignes sur un Réseau de Bus. ROADEF 2023 : 24 ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision., Rennes, France.

Titre : Construction d'une méthodologie et d'un outil associé pour la conception et la planification d'un réseau de bus

Mot clés : Conception de réseau de bus, Conception de lignes, Détermination de fréquences, Affectation de passagers, Génération de colonnes, Niveau de service, Mobilité, Programmation bi-niveaux

Résumé : L'entreprise Lumiplan propose actuellement le logiciel Heurès aux entreprises de transport public pour soutenir leurs opérations de planification des services matériels et conducteurs. Basé sur un partenariat entre IMT Atlantique et la société Lumiplan, ces travaux s'inscrivent dans une volonté de l'entreprise Lumiplan de proposer, en complément du logiciel Heurès, un outil d'optimisation des réseaux de bus au niveau des étapes de conception de lignes et de détermination des fréquences. Cette thèse aborde les deux premières étapes du processus global de planification des bus : le problème de conception des lignes et la détermination des fréquences. Le problème de conception des lignes vise à concevoir le réseau de bus pour une large période de temps tandis que le problème de détermination des fréquences vise à définir les fréquences des lignes de bus du réseau pour différentes périodes de temps d'une journée. L'étape de détermination des fréquences complète donc la conception des lignes avec des estimations plus précises de la demande, des types de bus nécessaires et du comportement des passagers. Pour l'étape de conception des lignes, nous étendons un modèle existant de la littérature avec des caractéristiques

réalistes comme les intervalles de fréquence des lignes et le nombre maximum de lignes, utiles pour les sociétés de transport public. Le problème est résolu par un processus innovant combinant des méthodes de génération et d'énumération de colonnes. Nous minimisons les coûts kilométriques tout en respectant de nouveaux paramètres de niveau de service pour les passagers. De plus, une comparaison entre nos résultats pour le problème de conception de lignes et ceux de la littérature sur deux instances existantes est effectuée. Pour l'étape de détermination des fréquences, nous minimisons les coûts d'exploitation de l'opérateur tout en respectant les contraintes de niveau de service, en se basant sur les prédictions du choix du trajet effectué par les passagers. Le problème de détermination des fréquences est résolu par un processus facilement implémentable et une étude de cas basée sur un réseau réel est présentée pour montrer l'efficacité de notre méthode. Enfin, une étude de cas basée sur un réseau réel est également réalisée pour prouver l'efficacité de nos méthodes. Quelques expériences sont donc réalisées pour comparer les réseaux créés et les fréquences obtenues pour différentes valeurs de paramètres.

Title: Construction of a methodology and associated tool for the design and planning of a bus network

Keywords: Bus Network Design, Line Planning, Frequency Setting, Passenger Assignment, Column Generation, Service-level, Mobility, Bi-level programming

Abstract: Lumiplan currently offers Heurès software to public transport companies to support their vehicle and driver scheduling operations. Based on a partnership between IMT Atlantique and Lumiplan, this work is part of Lumiplan's desire to offer, in addition to the Heurès software, a tool for optimizing bus networks at the network design and frequency setting stages. This thesis addresses the two first steps of the global process of bus planning: the line planning problem and the frequency setting. The line planning problem aims at designing the bus network for a large time period while the frequency setting problem aims at defining the frequencies of the bus lines of the network for different time periods of a day. The frequency setting step hence complements the line planning with more accurate estimations of the demand, necessary bus types and passengers behaviors. For the line planning step, we extend an existing model from the literature with realistic characteristics like line frequency intervals and maximum

number of lines, useful for public transportation companies. The problem is solved by an innovative process combining column generation and column enumeration methods. We minimize the kilometric costs while respecting new service-level parameters for passengers. Furthermore, a comparison between our results for the line planning problem and those of the literature on two existing instances is made. For the frequency setting step, we minimize the operator's exploitation costs while respecting service-levels constraints, based on the predictions of the path choice made by the passengers. The frequency setting problem is solved by an easily implementable process and a case study based on a real network is presented to show the efficiency of our method. Finally, a case study based on a real network is also performed to prove the efficiency of our methods. Some experiments are hence carried out to compare the created networks and the obtained frequencies for different parameter values.