

UNIVERSITÉ DE MONTRÉAL

MÉTHODES DE DÉCOMPOSITION POUR LA PLANIFICATION INTÉGRÉE
DES ITINÉRAIRES D'AVIONS ET DES HORAIRES D'ÉQUIPAGES

ANNE MERCIER
DÉPARTEMENT DE MATHÉMATIQUES
ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIAE DOCTOR (Ph.D.)
(MATHÉMATIQUES DE L'INGÉNIEUR)
AOÛT 2006



Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-20832-8

Our file Notre référence

ISBN: 978-0-494-20832-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

MÉTHODES DE DÉCOMPOSITION POUR LA PLANIFICATION INTÉGRÉE
DES ITINÉRAIRES D'AVIONS ET DES HORAIRES D'ÉQUIPAGES

présentée par : MERCIER Anne

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de:

M. AUDET Charles, Ph.D., président

M. SOUMIS François, Ph.D., membre et directeur de recherche

M. GENDREAU Michel, Ph.D., membre

M. VANDERBECK François, Ph.D., membre externe

À mes filles, Marie et Béatrice

REMERCIEMENTS

Je remercie très sincèrement mon directeur de recherche, M. François Soumis, pour son support. Sa disponibilité, ses connaissances et son jugement m'ont beaucoup aidée. Je le remercie également pour sa confiance. Cela m'a permis de concilier efficacement et agréablement mon travail et ma famille.

Je remercie M. Charles Audet d'avoir présidé le jury, ainsi que MM. Michel Gendreau et François Vanderbeck d'avoir accepté d'être membres du jury.

Je remercie également M. Yvan Dumas, de AD OPT Technologies, de m'avoir fourni des données provenant de deux grandes compagnies aériennes.

Le travail présenté dans cette thèse a bénéficié d'une bourse d'étude du Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR).

RÉSUMÉ

Le problème intégré de la création des itinéraires d'avions et des horaires d'équipages consiste à déterminer un ensemble de routes d'avions et de rotations d'équipages tels que chaque vol est couvert par un appareil et un équipage, tout en satisfaisant un ensemble de contraintes supplémentaires. Alors que certaines contraintes supplémentaires ne concernent que les équipages ou les avions, des contraintes liantes imposent des temps minimums de connexion pour les équipages qui dépendent des connexions utilisées par les appareils. Plus précisément, une connexion entre deux segments de vol qui n'est pas suffisamment longue pour être suivie par un équipage lorsque les deux vols sont effectués par des appareils différents est une connexion dite *courte*. Le problème intégré des avions et des équipages doit donc inclure un ensemble de contraintes supplémentaires pour s'assurer qu'un équipage ne puisse faire une connexion courte que si un avion la fait aussi.

Dans la première partie de cette thèse, après avoir défini le problème en détail, nous présentons une revue de la littérature concernant l'emploi de modèles d'optimisation pour traiter individuellement l'affectation des types d'appareils aux vols, la création des itinéraires d'avions et celle des rotations d'équipages. Nous discutons ensuite de modèles pour traiter de façon intégrée certaines étapes de la planification en transport aérien.

Dans la seconde partie de la thèse, nous proposons un modèle incluant un élément de robustesse pour traiter les contraintes liantes de connexions courtes et nous comparons deux méthodes de décomposition de Benders - une avec le problème des itinéraires d'avions dans le problème-maître et l'autre avec le problème des rotations d'équipages. Une solution est considérée plus robuste quand des perturbations à l'horaire sont moins susceptibles de retarder les vols suivant l'incident. Nous étudions aussi l'impact

de la génération de coupes Pareto-optimales sur la vitesse de convergence de ces méthodes. Sur un ensemble d'instances basées sur des données réelles fournies par deux compagnies aériennes majeures, l'approche avec le problème des rotations d'équipages dans le problème-maître donne de bonnes solutions en des temps de calcul raisonnables et est clairement supérieure à celle où le problème des itinéraires d'avions est dans le problème-maître.

Dans la troisième partie, nous proposons une approche pour améliorer la qualité globale des solutions d'avions et d'équipages en considérant un niveau supplémentaire d'intégration permettant de prendre certaines décisions par rapport à l'horaire des vols. Le problème intégré de la création d'itinéraires d'avions et d'horaires d'équipages avec des modifications à l'horaire consiste à déterminer un ensemble de routes d'avions et de rotations d'équipages et à choisir une heure de départ pour chaque segment de vol à l'intérieur d'une fenêtre de temps donnée. Des contraintes liantes assurent que le même horaire est choisi à la fois pour les itinéraires d'avions et les rotations d'équipages et imposent des temps minimaux de connexion pour les équipages qui dépendent des connexions utilisées par les appareils ainsi que des heures de départ choisies. Nous proposons une formulation compacte du problème ainsi qu'une méthode de décomposition de Benders, comprenant une procédure de génération dynamique de contraintes, pour résoudre le problème de façon efficace. Pour un ensemble d'instances basées sur des données réelles fournies par deux compagnies aériennes, la flexibilité par rapport à l'horaire des vols dans un modèle intégré avions-équipages procure des économies importantes de coûts tout en assurant la réalisabilité des itinéraires d'avions et des rotations d'équipages fournis par le modèle.

Lorsque la décomposition de Benders est utilisée, avec en problème-maître la création des rotations d'équipages, le sous-problème de Benders ne génère, en fait, que des coupes de réalisabilité qui sont ajoutées au problème-maître à chaque itération de l'algorithme. Pour résoudre le problème intégré d'avions et d'équipages, il a aussi

été proposé dans la littérature de résoudre itérativement un problème de rotations d'équipages modifié, en ajoutant des coupes de réalisabilité provenant d'une famille de coupes prédéfinie, jusqu'à ce que l'ensemble des connexions courtes utilisé par les équipages dans la solution courante soit réalisable pour les avions. Dans la dernière partie de la thèse, nous présentons une comparaison théorique des coupes de Benders et des autres familles de coupes. Nous proposons aussi une méthode de renforcement simple qui peut être appliquée à tous les types de coupes de réalisabilité pour le problème intégré avions-équipages.

ABSTRACT

The integrated aircraft routing and crew scheduling problem consists in determining a minimum-cost set of aircraft routes and crew pairings such that each flight leg is covered by one aircraft and one crew, and side constraints are satisfied. While some side constraints involve only crews or aircraft, linking constraints impose minimum connection times for crews that depend on aircraft connections. More precisely, a connection that is not long enough to be made by a crew when the two associated legs are flown by different aircraft is said to be *short*. Hence, a model that integrates the aircraft routing and the crew pairing formulations must also include one linking constraint for each possible short connection.

In the first part of the thesis, after formally defining the problem, we present a review of the literature on optimization models to individually address fleet assignment, aircraft routing and crew scheduling. We then review some models that integrate some of the airline planning problems.

In the second part, we propose an enhanced model incorporating robustness to handle the short connection linking constraints and compare two Benders decomposition methods - one with the aircraft routing problem as the master problem and one with the crew pairing problem. A solution is said to be more robust when disruptions on the day of operations are less likely to delay the flights following the unforeseen event. We also study the impact of generating Pareto-optimal cuts on the speed of convergence of these methods. Computational experiments performed on test instances provided by two major airlines show that the proposed approach with the crew pairing problem as the Benders master problem yields high quality solutions in reasonable computing times and is clearly superior than the one with the aircraft routing problem as the master problem.

In the third part, we consider an additional level of integration by adding some flight scheduling decisions to the integrated aircraft routing and crew scheduling problem. In this integrated aircraft routing, crew scheduling and flight retiming problem, a minimum-cost set of aircraft routes and crew pairings must be constructed while choosing a departure time for each flight leg within a given time window. Here, linking constraints ensure that the same schedule is chosen for both the aircraft routes and the crew pairings, and impose minimum connection times for crews that depend on aircraft connections and departure times. We propose a compact formulation of the problem and a Benders decomposition method with a dynamic constraint generation procedure to solve it. Computational experiments performed on test instances provided by two major airlines show that allowing some flexibility on the departure times within an integrated model yields significant cost savings while ensuring the feasibility of the resulting aircraft routes and crew pairings.

When Benders decomposition is used with the crew pairing problem as the Benders master problem, the aircraft routing subproblem is, in fact, generating feasibility cuts that are added to the Benders crew pairing master problem. To solve the integrated aircraft routing and crew scheduling problem, one can also solve a constrained crew pairing problem iteratively, by adding feasibility cuts from a predefined family to the crew pairing problem until a solution is found where the short connection set used is maintenance feasible for the aircraft. In the last part of the thesis, we perform an in-depth theoretical comparison of different types of feasibility cuts for the integrated aircraft routing and crew scheduling problem. We also propose a simple procedure to strengthen these cuts. Computational experiments performed on test instances provided by two major airlines are presented to support the theoretical results.

TABLE DES MATIÈRES

DÉDICACE	iv
REMERCIEMENTS	v
RÉSUMÉ	vi
ABSTRACT :	ix
TABLE DES MATIÈRES	xi
LISTE DES FIGURES	xvii
LISTE DES TABLEAUX	xviii
INTRODUCTION	1
CHAPITRE 1 DESCRIPTION DU PROBLÈME	4
1.1 L'affectation des types d'appareils aux segments de vol	4
1.2 La création des itinéraires d'avions	6
1.3 La création des rotations d'équipages	8
1.4 L'affectation des rotations d'équipages au personnel	11

1.5	La planification tactique intégrée	12
CHAPITRE 2 REVUE DE LA LITTÉRATURE		14
2.1	L'affectation des types d'appareils aux segments de vol	14
2.2	La création des itinéraires d'avions	27
2.3	La création des rotations d'équipages	33
2.4	La planification tactique intégrée	45
2.4.1	Intégration partielle	45
2.4.2	Intégration complète	48
2.5	L'intégration des tournées de véhicules et des horaires d'équipages en transport en commun urbain	53
CHAPITRE 3 ORGANISATION DE LA THÈSE		57
CHAPITRE 4 A COMPUTATIONAL STUDY OF BENDERS DECOMPO- SITION FOR THE INTEGRATED AIRCRAFT ROUTING AND CREW SCHEDULING PROBLEM		61
4.1	Mathematical Formulation	74
4.1.1	Aircraft routing networks	75
4.1.2	Crew pairing networks	77

4.1.3	Model	78
4.2	Solution Methodology	82
4.2.1	Benders decomposition using the crew pairing problem as a master problem	83
4.2.1.1	Three-phase algorithm	87
4.2.2	Identifying strong cuts	90
4.2.3	Step-by-step description of the improved algorithm	93
4.3	Computational Experiments	95
4.3.1	Description of data sets	95
4.3.2	Summary of computational experiments	99
4.3.3	Comparisons with the extended crew pairing approach of Cohn and Barnhart (2003)	103
4.4	Conclusion	109
CHAPITRE 5	AN INTEGRATED AIRCRAFT ROUTING, CREW SCHE- DULING AND FLIGHT RETIMING MODEL	113
5.1	Mathematical Formulation	127
5.1.1	Model	127

5.1.2	A simpler formulation	131
5.2	Solution Methodology	133
5.2.1	Benders decomposition	133
5.2.1.1	Three-phase algorithm	137
5.2.1.2	Dynamic constraint generation	138
5.3	Computational Experiments	139
5.3.1	Description of data sets	141
5.3.2	Summary of computational experiments	142
5.4	Conclusion	148
CHAPITRE 6	A THEORETICAL COMPARISON OF FEASIBILITY CUTS FOR THE INTEGRATED AIRCRAFT ROUTING AND CREW PAIRING PROBLEM	152
6.1	Mathematical Formulation	160
6.2	Benders Decomposition	161
6.2.1	Extreme rays	164
6.3	Some Predefined Families of Feasibility Cuts	167

6.4	Bounding the Benders Dual Subproblem	172
6.4.1	Model (a): One artificial variable in each short connection linking constraint	173
6.4.2	Model (b): A unique artificial variable appearing in every short connection linking constraint	183
6.4.3	Model (c): A unique artificial variable appearing in the convex- ity constraint	186
6.4.4	Comparison of the different bounding approaches for the Ben- ders dual subproblem	186
6.4.4.1	Cuts with a right-hand-side of value zero	187
6.5	Strengthening the Feasibility Cuts	190
6.5.1	Pareto-optimal cuts	191
6.5.2	Lifting procedure	192
6.6	Computational Experiments	196
6.6.1	Description of data sets	196
6.6.2	Summary of computational experiments	197
6.7	Conclusion	204

CHAPITRE 7	DISCUSSION GÉNÉRALE ET CONCLUSION	210
7.1	Contributions scientifiques de la thèse	210
7.2	Retombées pour l'industrie du transport aérien	213
7.3	Travaux futurs	214
BIBLIOGRAPHIE		222

LISTE DES FIGURES

Figure 4.1	Waiting time cost function	98
Figure 5.1	Example of some violated unaggregated constraints	133
Figure 5.2	Algorithmic flowchart	140
Figure 6.1	Crew costs as a function of the number of iteration - ECP (320D and MD80LA)	208
Figure 6.2	Crew costs as a function of the number of iteration - ECP RA (320D and MD80LA)	208
Figure 6.3	Crew costs as a function of CPU time - ECP RA (320D and D9SB)	209

LISTE DES TABLEAUX

Tableau 2.1	Caractéristiques des principaux modèles d'affectation des types d'appareils	25
Tableau 2.2	Caractéristiques des principaux modèles de routage d'avions	33
Tableau 2.3	Caractéristiques des principaux modèles de rotations d'équipages	44
Tableau 2.4	Caractéristiques des principaux modèles intégrant des étapes de la planification	54
Tableau 4.1	Summary of notation	82
Tableau 4.2	Characteristics of test instances	96
Tableau 4.3	Phase III Solution quality comparisons (% average maximum optimality gap)	100
Tableau 4.4	Complete results: basic integrated model (IM) [†]	101
Tableau 4.5	Complete results: generalized integrated model (GIM) [†]	104
Tableau 4.6	Results for UM+ECP and improved Benders decomposition approaches [†]	107
Tableau 5.1	Summary of notation	129
Tableau 5.2	Characteristics of test instances	141

Tableau 5.3	Complete results: integrated problem with flight retiming [†] . .	144
Tableau 5.4	Computational effort and solution quality: with and without flight retiming [†]	147
Tableau 6.1	Characteristics of test instances	197
Tableau 6.2	Crew arcs corresponding to infeasible short connections are kept [†]	199
Tableau 6.3	Crew arcs corresponding to infeasible short connections are removed (RA) [†]	200
Tableau 6.4	Instance 757C - Results for the integer formulation [†]	202
Tableau 6.5	When the infeasible short connection arcs are kept - IMARF [†]	203
Tableau 6.6	Identifying the individually infeasible short connections - IMARF [†]	204

INTRODUCTION

Que ce soit pour le travail ou pour le plaisir, de plus en plus de gens utilisent le transport aérien. Ce mode de transport doit sa popularité à sa rapidité et à son confort pour les voyages sur de longues distances. De plus, suite à la déréglementation de la fin des années quatre-vingts aux États-Unis d'abord et au Canada ensuite, et la venue sur le marché de transporteurs à tarifs réduits, les prix sont devenus plus attrayants. Malgré les bouleversements externes des dernières années, soit les événements du 11 septembre 2001, la guerre en Irak et l'épidémie de SRAS, l'on prévoit une hausse de 33% du nombre de passagers d'ici 2014 (ATAC, 2005). Le nombre élevé de passagers ainsi que les développements technologiques des quinze dernières années sont un incitatif pour que les transporteurs rationalisent leurs opérations et développent de nouvelles approches automatisées pour la planification de leurs vols. Une compagnie aérienne est très coûteuse à opérer et les économies à faire à l'aide d'une bonne planification peuvent être très importantes.

La planification des opérations d'une compagnie de transport aérien de passagers est une activité très complexe. Plusieurs étapes sont nécessaires afin d'assurer le bon déroulement des opérations. On regroupe habituellement les décisions de planification en trois principaux niveaux. La planification *stratégique* consiste en des décisions ayant des implications durant plusieurs années telles l'acquisition d'appareils et les décisions d'aménagement des aéroports. Le niveau *tactique*, celui qui nous intéresse particulièrement, concerne la planification à moyen terme qui doit être révisée à tous les trois ou quatre mois selon l'évolution de la demande. Le niveau *opérationnel* touche finalement aux décisions de très court terme prises en considérant les changements de dernière minute.

L'horaire des vols est à la base de la planification tactique. Le transport des passagers, contrairement au transport des marchandises, est très sensible à l'horaire, ce qui contribue à rendre encore plus ardue l'élaboration de ce dernier. On doit tenir compte, entre autres, de la répartition géographique des passagers, de leur tolérance à l'attente, de l'horaire des autres compagnies aériennes et de la configuration des aéroports. L'horaire des vols est composé d'une liste de segments de vols sans escale. Chacun de ces segments de vol est défini par son heure de départ, son heure d'arrivée ainsi que les aéroports d'origine et de destination. L'horaire est habituellement fixé pour environ trois mois, mais peut être légèrement modifié d'un mois à l'autre. Les transporteurs séparent leurs horaires en deux grandes catégories : les vols locaux (ou de courte durée) et les vols internationaux (ou de longue durée). Suite à l'élaboration de l'horaire vient une série de décisions. Il y a en fait quatre autres problèmes tactiques bien définis : l'affectation d'un type d'appareil à chaque vol, la création des itinéraires d'avions, la création des rotations d'équipages et l'affectation des rotations d'équipages au personnel. Ces problèmes sont présents à la fois dans le cas des vols de courte et de longue durée, mais ne sont pas traités de la même manière, en partie à cause de la configuration du réseau qui diffère dans les deux catégories. Aux États-Unis, par exemple, le réseau des vols locaux d'un transporteur est généralement de type *hub-and-spoke*, où chaque vol passe par un des aéroports principaux de la compagnie (*hub*) pour aller ou revenir d'un aéroport secondaire (*spoke*). Dans le cas des vols internationaux, il peut y avoir une suite de vols, pour les appareils ou les équipages, avant de revenir à un des aéroports principaux.

Les différents problèmes de la planification tactique ont traditionnellement été résolus de manière séquentielle. En fait, il demeure impossible, à ce jour, de traiter simultanément tous ces problèmes en raison de la grande taille du modèle qui résulterait d'une intégration complète. Une approche séquentielle réduit considérablement la complexité de la planification, mais cela peut donner des résultats de moindre qua-

lité puisque les décisions prises à un niveau donné peuvent réduire considérablement l'ensemble des solutions réalisables aux niveaux subséquents. L'objectif principal de cette thèse est le développement de modèles et d'approches de résolution pour intégrer la création des itinéraires d'avions et celle des rotations d'équipages dans le but de parvenir à une meilleure utilisation globale des ressources, et ce, pour des problèmes de grande taille.

CHAPITRE 1

DESCRIPTION DU PROBLÈME

Ce premier chapitre est consacré à la description des différentes étapes de la planification tactique des opérations d'un transporteur aérien de passagers. Nous les présentons dans l'ordre dans lequel les problèmes sont habituellement résolus : l'affectation d'un type d'appareil à chaque vol, la création des itinéraires des avions, la création des rotations des équipages et l'affectation des rotations d'équipages au personnel. Nous terminons ce chapitre par la description de quelques avantages et difficultés de la planification intégrée de ces étapes.

1.1 L'affectation des types d'appareils aux segments de vol

Le premier problème est l'affectation d'un type d'appareil à chacun des segments de vol. Les compagnies aériennes possèdent généralement plusieurs types d'avions avec des caractéristiques différentes. Une différence majeure entre deux types d'appareils est évidemment la capacité de l'avion, mais il y a aussi la vitesse, la consommation de carburant, le poids à l'atterrissage (qui détermine les frais d'aéroport), la portée, les coûts d'entretien, le temps de préparation requis entre deux vols, le niveau de bruit et la présence d'équipement de vol spécifique (par exemple pour voler au-dessus de l'eau). Lorsque l'on affecte un type d'appareil à un segment de vol, il faut éviter de choisir un avion de trop petite capacité qui risquerait de faire perdre des revenus potentiels à la compagnie aérienne en laissant des passagers au sol (*spill*), ou un avion de trop grande capacité, dont les coûts d'opération seraient trop élevés si la demande ne permettait pas de le remplir.

Une des principales difficultés de la formulation de ce problème est l'estimation des bénéfices spécifiques aux segments de vol par type d'appareil. Alors que les coûts d'opération sont assez faciles à déterminer, les revenus provenant des passagers sont difficiles à estimer. Ils dépendent des prévisions de la demande, dont une partie ne pourra être satisfaite faute de place dans les appareils. Un certain nombre de ces passagers laissés de côté pourront être replacés sur un autre vol de la compagnie, alors que les autres seront perdus au profit d'une autre compagnie aérienne. Comme les compagnies aériennes n'ont pas d'information sur cette dernière portion, elle est très difficile à estimer. De plus, les segments de vols ne sont pas indépendants les uns des autres. Par exemple, si un grand nombre de vols de Montréal à Vancouver passent par Toronto, alors la demande non satisfaite de Montréal à Toronto peut influencer la demande entre Toronto et Vancouver. Posséder de l'information sur la demande des passagers en fonction de l'origine vers la destination finale est important pour pouvoir bien estimer les coûts de la demande non satisfaite, mais l'estimation de la demande de l'origine vers la destination finale est en elle-même un problème d'optimisation.

Une séquence de segments de vol peut porter le même numéro de vol si le même appareil couvre tous les segments de la séquence. Nous appellerons ce type de séquence une séquence combinée (*thru*). Par exemple, un vol de Montréal à Vancouver en passant par Toronto comprend deux segments de vol qui pourraient porter le même numéro. Les compagnies aériennes choisissent généralement les séquences combinées à l'étape de l'affectation des types d'appareils. Les passagers qui voyagent sur deux segments d'une séquence combinée peuvent demeurer à bord de l'appareil, évitant ainsi de transiter par l'aéroport pour effectuer leur correspondance. Ces paires de vols sont intéressantes pour les passagers car elles leur assurent de ne pas manquer leur connexion tout en évitant une manipulation supplémentaire des bagages. Comme les gens sont habituellement prêts à payer un supplément pour ce genre de voyage, les compagnies ont intérêt à bien choisir ces paires. Les séquences combinées restreignent

par contre les possibilités lors de la planification car l'ensemble des vols formant une séquence combinée doit nécessairement être affecté non seulement au même type d'avion mais au même appareil. Par ailleurs, il y a certaines contraintes à respecter lors du choix de ces séquences. Par exemple, la patience des passagers dicte une borne supérieure sur le temps de connexion entre deux segments successifs d'une séquence combinée et seules les séquences qui n'augmentent pas radicalement le temps de vol par rapport à un vol direct sont possibles (par exemple, un vol de Montréal à Québec en passant par Toronto ne serait pas acceptable pour les passagers). On choisit alors les séquences combinées parmi celles qui sont possibles de façon à maximiser le revenu additionnel en tenant compte de leur effet sur l'affectation des appareils à tous les segments de vol.

Il y a trois groupes de contraintes que tout modèle d'affectation des types d'appareils doit respecter : chaque segment de vol doit être couvert par exactement un type d'appareil, le nombre d'avions entrant et sortant de chaque aéroport doit être équilibré et l'on doit respecter le nombre d'avions disponibles de chacun des types. Dans le cas de la planification stratégique, cette dernière contrainte pourrait néanmoins être absente de manière à justement déterminer la composition de la flotte d'appareils de la compagnie. La plupart des transporteurs doivent par ailleurs respecter certaines contraintes spéciales, telles des restrictions à certaines heures pour les appareils dont le niveau de bruit est trop élevé ou des limites sur la capacité de certaines portes d'embarquement à l'aéroport d'origine ou de destination.

1.2 La création des itinéraires d'avions

Le deuxième problème est la création des itinéraires des appareils. La solution du problème d'affectation des types d'appareils aux segments de vol fournit une

affectation respectant le nombre d'appareils disponibles de chacun des types. Il faut ensuite déterminer l'itinéraire de chaque appareil en s'assurant qu'il permette l'entretien de celui-ci. Chaque compagnie aérienne respecte des normes de sécurité dictées par les agences gouvernementales. De plus, les compagnies aériennes ont souvent leur propre ensemble de règles, généralement plus strictes que celles des gouvernements. Il y a plusieurs catégories d'entretien qui se distinguent par l'ampleur de la vérification et l'intervalle de temps entre les séances. Les entretiens qui ont lieu à tous les six mois ou à chaque année ne font habituellement pas partie de la planification tactique, qui est établie pour un horizon d'environ trois mois. Pour s'assurer de respecter ces séances d'entretien d'envergure, on réduit tout simplement le nombre d'avions disponibles selon le nombre d'avions devant se rendre en même temps à ce type d'entretien. On doit toutefois planifier explicitement les entretiens réguliers qui se font à tous les trois ou quatre jours (ou après une soixantaine d'heures de vol). Ces entretiens, qui sont des vérifications visuelles de différentes pièces de l'appareil, prennent environ six heures et nécessitent la présence d'un personnel qualifié à l'aéroport. C'est pourquoi seulement quelques aéroports peuvent effectuer l'entretien des appareils. La localisation des stations d'entretien peut aussi dépendre du type d'appareil. Dans le cas d'un horaire local, où les vols sont généralement planifiés pour avoir lieu le jour, les avions sont inactifs la nuit. Il y a donc beaucoup d'opportunités pour des séances d'entretien nocturnes et elles deviennent faciles à planifier. Le problème de la création des itinéraires d'avions est dans ce cas un problème facile à résoudre, contrairement aux itinéraires d'avions réalisant les vols de longue durée.

On peut créer les itinéraires des avions de façon à ce qu'ils respectent aussi les entretiens cycliques qui ont lieu à tous les trois mois. Ces entretiens ne peuvent être effectués qu'à un sous-ensemble des stations d'entretien régulier et prennent une douzaine d'heures. Alors qu'on n'impose habituellement pas de limite sur le nombre

d'appareils qui peuvent subir un entretien régulier en même temps, il y a une limite de capacité à respecter dans le cas des entretiens cycliques. Le problème de la création des itinéraires d'avions avec l'entretien cyclique est plus difficile à résoudre, même pour les vols de courte durée.

1.3 La création des rotations d'équipages

Le troisième problème de la planification tactique est la création des rotations d'équipages. Parmi les coûts d'opération, les coûts d'équipages se classent deuxièmes en importance derrière les coûts de carburant. Il est donc important pour les compagnies aériennes d'optimiser les horaires de leur personnel. Avec les années, on a remarqué que non seulement les compagnies pouvaient réduire considérablement leurs coûts d'opération grâce à l'optimisation, mais les rotations construites de cette façon augmentaient de beaucoup la qualité des horaires du point de vue des employés. Le problème des rotations d'équipages consiste à construire un ensemble de rotations, par famille d'appareils, de façon à couvrir tous les vols de la famille tout en respectant un ensemble de règles dictées par le contrat de travail des employés. Une famille d'appareils est un ensemble de types d'appareils qui partagent les mêmes exigences en ce qui concerne l'équipage. Le personnel qualifié pour travailler avec un type d'appareil de la famille peut travailler avec tous les types de la famille.

Chaque employé est affecté à une des bases de la compagnie et les rotations commencent et se terminent à une de ces bases. Les bases des équipages ne sont pas nécessairement les mêmes que les stations d'entretien, mais dans le cas des vols de courte durée, aux États-Unis, les aéroports principaux sont habituellement à la fois des stations d'entretien et des bases d'équipage. Une rotation est formée d'une ou plusieurs journées de travail séparées par des repos de nuit et chaque journée de

travail est composée d'un ou plusieurs segments de vol séparés par des connexions. Plusieurs éléments de la rotation sont contraints par le contrat de travail des employés. Par exemple, pour éviter que les employés demeurent très longtemps hors de leur base, une rotation doit généralement durer au maximum quatre ou cinq jours dans le cas des vols locaux, alors qu'elle peut durer jusqu'à une dizaine de jours pour les vols internationaux. Il y a aussi une durée minimale et une durée maximale pour les temps de connexion. Comme il faut un temps minimal pour que l'équipage prépare le vol suivant mais que si l'attente est trop longue, cette attente devient un repos afin que l'équipage puisse aller se reposer à l'hôtel, les temps de connexion sont généralement compris entre 45 minutes et quatre heures. Les repos ont quant à eux une durée minimale d'environ dix heures. Certaines règles sont toutefois plus complexes. Par exemple, un repos pourrait être plus court que les dix heures réglementaires si l'équipage a eu un repos de longue durée la veille ou doit en avoir un le lendemain. Finalement, la durée totale d'une journée de travail est contrainte à être d'au plus une dizaine d'heures, et le nombre d'heures de vol et le nombre de segments de vol par journée de travail sont aussi contraints à être d'au plus huit heures et quatre segments, respectivement.

Le temps de connexion minimal entre deux vols pour un équipage peut dépendre des itinéraires d'avions. Plus précisément, le temps minimal de connexion peut être plus court si l'équipage demeure à bord du même avion. On dit alors que l'équipage suit l'avion et le temps de préparation pour le prochain vol est réduit. Réduire le temps minimal de connexion de 45 minutes à 30 minutes, par exemple, peut permettre des connexions qui étaient impossibles et diminuer ainsi le coût des rotations.

Le contrat de travail garantit aussi généralement un salaire minimal par rotation. C'est-à-dire que, comme il n'y a pas de durée totale minimale par rotation, un certain nombre d'heures non travaillées peuvent être payées. La structure de coûts est donc relativement complexe et est le maximum de deux éléments : a) une fraction de la

durée totale de la rotation multipliée par une constante (car les attentes sont moins payées que le temps de vol) et b) la somme des coûts des journées de travail. Le coût d'une journée de travail est aussi généralement le maximum de trois éléments : a) un minimum garanti, b) une fraction de la durée totale de la journée multipliée par une constante et c) le coût réel de la journée de travail. Le coût réel de la journée de travail dépend du nombre d'heures de vol, de la durée des connexions, du coût des repos (hôtel, transport, nourriture) et du coût du transport de repositionnement. Le transport de repositionnement peut être soit terrestre lorsque l'équipage doit prendre un autobus ou un taxi pour changer d'aéroport (par exemple de La Guardia à J.F. Kennedy à New York), soit aérien lorsque les membres de l'équipage deviennent passagers sur un vol qui leur permet de se rendre à l'aéroport d'origine de leur prochain segment de vol ou de se rendre à leur base à la fin de la rotation. Les vols de repositionnement (*deadheads*) permettent souvent une réduction considérable des coûts d'équipages, bien qu'un coût leur soit rattaché. Si l'équipage utilise un vol de la compagnie, alors le coût est le revenu perdu des sièges occupés par l'équipage ; si le vol en est un d'une autre compagnie, alors le coût est celui du billet d'avion. Une fraction de leur salaire horaire peut aussi être versée aux employés en transit.

Le coût d'une rotation peut aussi être donné sous forme de *crédits de vol*. La différence entre le coût de la rotation et le coût réel de la partie comptant uniquement le temps de vol est appelée le coût en excès et les crédits sont le rapport entre les coûts en excès et le coût relié au temps de vol. Les compagnies aériennes désirent évidemment que les crédits soient les plus petits possible et cette façon de calculer les coûts leur permet de voir rapidement l'ampleur des réductions de coûts possibles.

Les contraintes décrites jusqu'ici sont des contraintes locales, régissant les rotations ou les journées de travail, mais il existe aussi des contraintes globales régissant l'ensemble des rotations. Comme chaque employé est affecté à une base, les rotations doivent ensemble respecter le nombre d'employés disponibles à chacune des bases. De plus,

les compagnies aériennes désirent souvent distribuer le plus équitablement possible le nombre d'heures de travail à travers les différentes bases. D'autres contraintes globales, comme un nombre maximal de vols de repositionnement ou de changements d'avion peuvent aussi être présentes.

Il y a en fait deux problèmes distincts de rotations d'équipages : un pour les pilotes et un autre pour les agents de bords. Ceci s'explique par le fait que les règles du contrat de travail ainsi que les familles d'appareils sont différentes. Les pilotes sont en effet qualifiés pour un poste de pilotage précis, alors que les agents de bord peuvent travailler à bord d'un éventail plus large de types d'avions.

1.4 L'affectation des rotations d'équipages au personnel

La quatrième et dernière étape de la planification est l'affectation des rotations d'équipages au personnel afin de créer les horaires mensuels de chacun des employés. Ces horaires doivent tenir compte d'un ensemble de règles définies par le contrat de travail, comme un nombre de jours de congé minimal entre les rotations. Il existe trois façons bien définies de faire ces horaires. Premièrement, on peut créer le nombre d'horaires voulus et laisser ensuite chaque employé choisir selon son ancienneté (*bidding*). Une autre méthode consiste à créer des horaires personnalisés pour chaque employé en maximisant une mesure de la satisfaction globale des employés (*rostering*). Finalement, on peut aussi faire des horaires personnalisés, mais qui satisfont les préférences des employés selon leur ancienneté (*preferential bidding system*).

1.5 La planification tactique intégrée

Les problèmes décrits précédemment peuvent être individuellement difficiles à résoudre en raison de leur grande taille. Une compagnie aérienne de taille moyenne comme Air Canada compte environ 1000 vols par jour et ce nombre peut être multiplié par trois pour certaines grandes compagnies aériennes américaines. Utiliser une approche séquentielle réduit considérablement la complexité de la planification. Par exemple, affecter les types d'avions avant le routage permet de séparer le problème du routage par type d'appareil. Bien sûr, cela donne plusieurs problèmes de routage, mais chacun devient beaucoup plus facile à résoudre. Par contre, la perte flexibilité résultant d'une procédure séquentielle peut donner des résultats de moindre qualité. Les décisions prises à un niveau donné peuvent même rendre non réalisables les problèmes des niveaux subséquents. Par exemple, après avoir affecté les types d'appareils aux segments de vol, il pourrait être impossible, pour un type d'appareil donné, de trouver des itinéraires respectant les contraintes d'entretien alors qu'il en existerait avec une autre affectation des types d'appareils.

Depuis quelques années, on observe un intérêt croissant pour l'intégration de la création des itinéraires d'avions et celle des rotations d'équipages. CORDEAU *et al.* (2001), KLABJAN *et al.* (2002) et COHN et BARNHART (2003) ont montré que la qualité des solutions obtenues avec l'intégration est significativement supérieure à la qualité des solutions produites par la résolution séquentielle. La principale interaction entre ces deux étapes est un temps de connexion variable. Comme le temps minimal de connexion entre deux vols couverts par le même équipage est réduit si le même avion couvre aussi les deux vols (l'équipage, en restant dans le même avion, n'a pas à se déplacer dans l'aéroport), l'ensemble des rotations d'équipages réalisables dépend des itinéraires d'avions construits. Les coûts variables d'équipages provenant en grande

partie des repos et des connexions, une solution sous-optimale risque d'être trouvée par un processus séquentiel.

Lorsque le problème des avions et celui des équipages sont résolus séquentiellement, on commence habituellement par fabriquer les itinéraires des avions. Selon les itinéraires choisis pour les avions à la première étape, on peut déterminer l'ensemble des connexions possibles pour les équipages et ensuite s'attaquer à la résolution de la deuxième étape. Une connexion entre deux segments de vol qui n'est pas assez longue pour pouvoir être suivie par un même équipage lorsque les deux vols sont faits par des appareils différents est une connexion dite *courte*. On peut facilement traiter le lien entre les deux problèmes implicitement dans la structure du réseau des équipages car l'ensemble des connexions courtes possibles pour les équipages est fixé par la solution du problème des avions, qui n'est résolu qu'une seule fois. Ceci est fait en omettant les arcs qui correspondent aux connexions courtes qui ne sont pas suivies par un avion. On s'assure ainsi de respecter le temps minimal de connexion. Par contre, lorsque l'on formule le problème intégré des avions et des équipages, on doit inclure un ensemble de contraintes supplémentaires pour s'assurer qu'un équipage ne puisse faire une connexion courte que si un avion la fait aussi. Ceci est nécessaire car il n'y a plus d'étape où l'on fixe les connexions. Le modèle intégré, plus gros que l'union des deux modèles individuels, est donc beaucoup plus difficile à résoudre.

Étant donné les coûts importants reliés à l'utilisation d'un avion, un modèle plus étendu intégrant aussi le choix du type d'appareil affecté à chaque segment de vol permettrait des économies encore plus considérables pour les transporteurs aériens. La difficulté principale de ce dernier problème provient du très grand nombre de contraintes liant les trois étapes. Ces contraintes sont nécessaires afin de s'assurer que l'équipage affecté à un vol possède des qualifications compatibles avec le type d'appareil affecté à ce même vol.

CHAPITRE 2

REVUE DE LA LITTÉRATURE

En recherche opérationnelle, la littérature contient peu de références au sujet de l'intégration des niveaux de la planification tactique en transport aérien, mais plusieurs travaux ont été faits sur l'optimisation des problèmes individuels. Par ailleurs, certains auteurs, sans intégrer totalement les niveaux de décision, améliorent l'optimisation d'un problème en considérant certains aspects des autres problèmes.

Ce chapitre présente premièrement une revue détaillée de la littérature concernant l'emploi de modèles d'optimisation pour traiter individuellement l'affectation des types d'appareils, la création des itinéraires d'avions et celle des rotations d'équipages. Une section est ensuite consacrée aux travaux qui visent l'intégration de certaines de ces étapes de la planification. Finalement, une brève revue de la littérature concernant le problème intégrant la fabrication des tournées de véhicules et les horaires d'équipages en transport en commun urbain est incluse.

2.1 L'affectation des types d'appareils aux segments de vol

Le problème de l'affectation des types d'appareils aux segments de vol revêt une grande importance pour les compagnies aériennes car les coûts qui y sont rattachés sont très élevés. Plusieurs auteurs se sont donc intéressés à ce problème dont la complexité a été étudiée par GU *et al.* (1994).

Les horaires de vols comportent habituellement des groupes de segments qui se répètent de façon quotidienne ou hebdomadaire. Certains vols ont lieu chaque jour, ou tous les lundis, par exemple. Cet aspect cyclique des horaires réduit la complexité des problèmes et en facilite la résolution. Les compagnies aériennes ne traitent donc habituellement pas les problèmes sur toute la période de planification, mais sur un jour ou une semaine. L'horaire des vols locaux reste sensiblement le même durant la semaine et comporte une quantité réduite de vols la fin de semaine. Pour les vols locaux, on résout généralement un problème sur une journée en considérant les segments qui sont présents au moins quatre fois durant la semaine et des ajustements sont ensuite faits pour se conformer à l'horaire réel. Les vols internationaux sont quant à eux souvent résolus sur une semaine.

Ce problème d'optimisation est généralement formulé comme un problème de réseau multi-flots en variables mixtes, où les différents flots sont associés aux types d'appareils. Des contraintes liants les types d'appareils sont présentes pour s'assurer de la couverture de tous les vols. Ce sont ces contraintes ainsi que certaines autres contraintes supplémentaires (souvent des restrictions sur l'utilisation de certains types d'avions) qui détruisent la structure de réseau pur. On le résout donc avec une méthode de séparation et d'évaluation progressive (*branch-and-bound*).

Le réseau sous-jacent a une structure *espace-temps*. Une manière de le définir est de placer les segments de vol sur les arcs. Les noeuds représentent alors l'origine ou la destination d'un segment de vol. Outre les arcs de vol, d'autres arcs sont présents pour représenter les connexions possibles entre les segments. Cette façon traditionnelle de définir le réseau fait en sorte qu'il croît très rapidement avec le nombre de vols. Certains auteurs qui utilisent ce type de réseau tentent de réduire sa taille en choisissant arbitrairement un sous-ensemble de connexions à mettre dans le réseau. On peut aussi faire des modifications au réseau traditionnel pour le rendre plus compact, sans laisser de côté certaines possibilités. Un réseau qu'on appelle à

lignes de temps (time line) permet de le faire. Les noeuds représentent toujours le départ ou l'arrivée des vols, mais en les mettant cette fois en ordre chronologique, on peut éliminer les arcs de connexion en introduisant des arcs d'attente ne reliant une arrivée qu'au départ suivant. Tous les avions qui sont en attente dans la même ville à un moment donné empruntent alors le même arc. Il y a une ligne de temps par ville et le nombre d'arcs est ainsi diminué significativement, sans perdre de généralité. Ce type de réseau, beaucoup plus efficient que la structure traditionnelle, est très utilisé dans les modèles d'affectation des types d'appareils.

La fonction-objectif tient compte des coûts associés à la demande insatisfaite (les revenus perdus en raison du manque de capacité) et des coûts d'opération qui dépendent à la fois du type d'avion et du segment de vol. Les coûts d'opération sont les coûts réels, donc faciles à obtenir, et les revenus perdus sont estimés plus ou moins rigoureusement selon les auteurs. Selon BELOBABA et FARKAS (1999), même si l'on fait des progrès pour pouvoir résoudre des problèmes de grande taille, on continue à utiliser une approche agrégée pour estimer les revenus perdus, et ce, même si l'on doit considérer plusieurs classes de tarifs. On peut agréger horizontalement la demande de tous les départs sur la période de planification. En faisant cela, on suppose que la demande à l'intérieur de la période de planification est homogène et on représente la demande pour un certain segment de vol pendant une période comme une distribution de probabilités gaussienne simple. En pratique, les compagnies aériennes génèrent cette simple demande moyenne pour un segment de vol en regroupant un échantillon d'observations historiques sur les départs de la période considérée pour faire une prévision de la demande future pour ce segment. On peut aussi agréger verticalement la demande en combinant les demandes de plusieurs classes de tarif en une seule fonction de densité pour un seul départ d'un segment de vol. Cette méthode ne tient pas compte des différences de demande entre les classes de tarif. Toujours selon les auteurs, l'agrégation verticale peut compenser, en théorie, pour l'agrégation

horizontale, mais elles peuvent aussi s'additionner et ainsi donner un biais très grand qui peut même modifier l'affectation des appareils. Leurs tests montrent jusqu'à 33% de sous-estimation des coûts. Ils suggèrent une méthode désagrégée en trois étapes. Ils regroupent premièrement les départs qui ont les mêmes patrons de demande (par exemple, les lundis). Le système de réservation de la compagnie, à l'aide de son module de gestion des revenus (*yield management*), peut ensuite générer des prévisions de coûts par départ et par classe de tarif. Cette deuxième étape n'est pas utile pour les problèmes sur une journée car le même type d'appareil doit être affecté à un même segment de vol chaque jour. Ils estiment donc ensuite les coûts pour toute la période de planification comme une moyenne pondérée pour chaque groupe défini à la première étape à l'aide des prévisions fournies par la deuxième étape.

Tous les auteurs imposent les trois groupes de contraintes suivants : la couverture des vols, la conservation de l'équipement à chaque noeud du réseau (conservation du flot) et le respect du nombre d'avions de la compagnie disponibles pour chaque type d'appareil.

ABARA (1989) présente un modèle linéaire en nombres entiers à partir d'un réseau traditionnel. L'auteur restreint le nombre d'arcs de connexion en connectant chaque segment de vol à au plus cinq autres segments. Cette réduction à seulement cinq connexions limite beaucoup le modèle, surtout si le réseau est de type *hub-and-spoke*, où il peut exister des intervalles de temps de courte durée, appelés complexes, pouvant comprendre jusqu'à 65 arrivées et le même nombre de départs. De plus, le modèle ne tient pas compte de la possibilité que les segments de vol aient une durée et un temps de préparation qui soient différents selon le type d'avion. Le modèle comprend par ailleurs plusieurs contraintes spéciales qui en augmentent le réalisme. On peut, par exemple, mettre une limite sur le nombre d'avions qui passent la nuit à certains aéroports. Plusieurs objectifs sont proposés, tels que la minimisation des

coûts d'opération ou la maximisation de l'utilisation des types d'appareils les plus efficaces.

HANE *et al.* (1995) présentent un modèle plus compact pour résoudre des problèmes domestiques de grande taille sur une journée. Le modèle se base sur un réseau à lignes de temps auquel on fait certaines agrégations de noeuds. Ces agrégations de noeuds regroupent les arrivées qui ne sont pas séparées par un départ et les départs qui ne sont pas séparés par une arrivée. Les auteurs introduisent aussi la notion d'îlots, qui exploitent la faible densité des vols aux aéroports secondaires. En enlevant certains arcs d'attente choisis, on crée des groupes de départs et d'arrivées isolés, des îlots, qui permettent d'éviter d'avoir des avions au sol inutilement. La structure d'îlots permet en plus de fixer certaines connexions s'il n'y a qu'une seule arrivée et un seul départ dans le groupe, ou d'en éliminer si l'on se rend compte qu'affecter un certain type d'appareil à un segment de vol ferait manquer une connexion subséquente. Un trop grand nombre d'îlots peut rendre le problème non réalisable ou donner des solutions loin de l'optimum. En choisissant des aéroports où il existe de très longs arcs d'attente, donc une structure naturelle d'îlots, on remarque une amélioration importante des temps de calcul sans détérioration marquée de la qualité des solutions. Outre ces manipulations de réseau, les auteurs suggèrent des méthodes de pré-traitement algébrique pour réduire la taille du problème. À chaque fois que le modèle est résolu, la valeur de certaines variables est fixée, selon diverses règles, dans le but de réduire la taille de l'arbre de branchement. Les auteurs décrivent aussi une méthode de séparation et d'évaluation qui équilibre les branches de l'arbre en fixant, d'un coup, un ensemble de variables à l'aide des contraintes de couverture. De plus, les contraintes de couverture forment des ensembles de variables que l'on peut ordonner selon une mesure de variabilité calculée à l'aide de l'information fournie par la fonction-objectif. L'idée est que l'on obtiendra plus d'impact si l'on impose une nouvelle contrainte sur un ensemble ayant une plus grande variabilité. On calcule la variabilité soit avec : a) la

différence entre le maximum et le minimum des coefficients dans l'objectif, b) la somme de la valeur absolue des différences des coefficients de l'objectif ou c) en estimant la dégradation de la valeur de l'objectif pour chacune des contraintes de branchement qui pourraient être ajoutées. Les auteurs suggèrent aussi d'affecter prioritairement un appareil à un vol de longue durée car cela retire un avion pour une plus longue période et pourrait donc avoir plus d'impact sur la relaxation linéaire du problème.

CLARKE *et al.* (1996) généralisent le modèle de HANE *et al.* (1995). Ils montrent qu'il est possible d'obtenir de meilleures solutions globales en intégrant au problème de l'affectation des types d'appareils aux vols quelques aspects du problème de routage des avions et du problème des rotations d'équipages. Des arcs supplémentaires sont ajoutés afin de créer des possibilités pour l'entretien régulier. Des limites sur le nombre d'heures de vol par famille d'appareils sont imposées à chaque base et des pénalités sont incluses dans l'objectif pour favoriser des départs de mi-journée suite à une arrivée en fin de soirée la veille, permettant ainsi de placer plus facilement par la suite les repos des équipages. Ces contraintes supplémentaires sont simples à inclure et facilitent la résolution des problèmes suivants, mais ne garantissent pas leur réalisabilité. Les auteurs reconnaissent que leur méthode ne permet pas d'inclure des possibilités pour les entretiens cycliques (aux trois mois). Ils utilisent les techniques d'agrégation et de résolution développées par HANE *et al.* (1995) et obtiennent des solutions réalisables pour certains problèmes de routage d'avions qui étaient non réalisables lorsque la solution d'affectation des types d'appareils ne considérait pas l'entretien.

SUBRAMANIAN *et al.* (1994) décrivent une application des méthodes proposées par CLARKE *et al.* (1996) et HANE *et al.* (1995) chez Delta Air Lines. Ils formulent donc le problème sur une journée, à partir d'un réseau à lignes de temps, et font les mêmes agrégations. Les auteurs estiment les coûts de revenus perdus à l'aide d'une courbe normale. Outre les contraintes habituelles, le modèle inclut les requêtes de séquences combinées. Le nombre maximal d'avions pour chacun des types peut

être dépassé, mais une grande pénalité est incluse dans l'objectif. Des contraintes supplémentaires qui rendent le modèle plus réaliste sont aussi présentes de même que des arcs supplémentaires avec pénalités pour permettre certaines connexions courtes intéressantes. Finalement, plusieurs objectifs peuvent être utilisés un à un, tels que minimiser les coûts ou minimiser le nombre d'avions. Le modèle peut aussi servir au choix des vols lors de la planification stratégique. Le modèle a été testé avec les données de Delta, comprenant plus de 2500 vols locaux chaque jour avec environ 450 avions de dix types différents. Les auteurs croient que leur modèle aura permis à la compagnie d'épargner quelques millions de dollars américains.

Un peu à la manière de ABARA (1989), RUSHMEIER et KONTOGIORGIS (1997) utilisent un réseau de type traditionnel pour un problème sur une journée. Ce type de réseau leur permet de distinguer chaque avion au sol. On peut ainsi empêcher certaines connexions non voulues par la compagnie aérienne ou mettre des coûts spécifiques de connexion, mais cela augmente de beaucoup la taille du réseau. Pour réduire le nombre d'arcs, les auteurs forment des groupes de noeuds, surtout aux aéroports secondaires, où seules les connexions dans un même groupe sont permises. Cette méthode généralise la structure d'îlots décrite par HANE *et al.* (1995). La fonction-objectif comprend donc, outre les coûts habituels, des possibilités de bonus ou de pénalités spécifiques aux connexions. Les auteurs incluent aussi des considérations pour l'entretien et les équipages en fixant des objectifs, par exemple, quant au nombre d'avions qui passent la nuit à une station d'entretien ou au nombre d'arrivées et de départs à une base d'équipage. Ces contraintes supplémentaires ne sont pas imposées de façon stricte, mais peuvent être violées moyennant des pénalités dans l'objectif. Une fonction concave linéaire par morceaux permet d'augmenter le coût lorsque la solution s'éloigne des objectifs fixés. Le pré-traitement et la méthode de résolution s'approchent de ceux utilisés par HANE *et al.* (1995), incluant aussi une méthode de séparation et d'évaluation qui s'appuie sur les bornes supérieures généralisées des

contraintes de couverture. Les auteurs ont testé leur méthode avec des données de US Air. Les problèmes comprenaient environ 1600 vols pour huit types d'appareils et leur modèle a permis à la compagnie de faire un bénéfice annuel de plusieurs millions de dollars américains. Par contre, la présence d'un nombre élevé de termes de pénalités rend la résolution ardue, surtout à cause des fonctions linéaires par morceaux qui sont modélisées à l'aide d'un ensemble de variables supplémentaires.

À l'aide d'un modèle d'affectation d'appareils aux segments de vol qui permet de modifier légèrement l'horaire de vol, certains auteurs améliorent les profits et réussissent à diminuer le nombre d'avions nécessaires à la couverture des vols. La méthode présentée par DESAULNIERS *et al.* (1997b) détermine l'horaire journalier qui maximise les profits dans le contexte d'une flotte hétérogène. Leur formulation, sans procurer un horaire cyclique, fait en sorte que le nombre d'appareils d'un même type disponibles à une base donnée au début d'une journée soit le même que le nombre d'appareils disponibles à cette base à la fin de la journée. Les auteurs présentent une méthode de partitionnement d'ensemble avec des contraintes supplémentaires où chaque variable représente un chemin dans un réseau. La relaxation linéaire est résolue par une méthode de génération de colonnes où des sous-problèmes de réseau génèrent des horaires prometteurs et réalisables par type d'appareil. Les réseaux incluent des arcs de connexions courtes avec une pénalité dans l'objectif. On peut minimiser le nombre d'avions en mettant une pénalité sur les arcs partant de la source et les séquences combinées sont modélisées à l'aide d'un seul arc. Comme il est plus difficile de faire un branchement avec des variables de chemin (on peut régénérer à une itération donnée un chemin mis à 0 à l'itération précédente), les auteurs utilisent plutôt les variables de flot retrouvées dans les sous-problèmes comme variables de branchement. Ils proposent des branchements heuristiques pour une combinaison de variables en imposant, par exemple, une connexion entre deux segments de vol, peu importe le type d'appareil. Les auteurs suggèrent une modification à la modélisation

dans le but de rendre leur modèle cyclique. Ils présentent des résultats pour deux ensembles de données : un transporteur nord-américain avec 383 vols par jour et 91 avions de neuf types ainsi qu'un transporteur européen de taille semblable. Trois scénarios, qui dépendent de la largeur permise pour les fenêtres de temps (soit 10, 20 ou 30 minutes), sont évalués pour chaque transporteur. On remarque une augmentation de la taille du réseau avec l'augmentation de la largeur des fenêtres de temps (plus de connexions sont possibles), mais aussi une diminution du nombre d'avions nécessaires. Les temps de calcul sont raisonnables, mais les problèmes ne sont pas de grande taille.

Lorsque le problème de l'affectation des types d'appareils aux vols se fait sur une semaine et que des modifications à l'horaire sont permises, il est habituellement préférable de synchroniser les vols qui ont lieu plus d'une fois pendant la semaine. IOACHIM *et al.* (1999) proposent une formulation et une méthode de résolution par génération de colonnes pour ce problème particulier. Le problème-maître consiste en un problème de partitionnement d'ensemble avec des contraintes supplémentaires de synchronisation et le sous-problème est un problème de plus court chemin avec des fenêtres de temps. Les auteurs proposent de nouvelles méthodes de branchement utilisant les variables d'horaire ainsi que certaines modifications dans le but d'améliorer la convergence de l'algorithme. Ces modifications incluent une réduction du nombre de contraintes dans le problème-maître suite à certaines décisions de branchement et l'introduction d'un paramètre de tolérance pour la synchronisation. Les auteurs ont testé leur méthode à l'aide de données provenant de la compagnie Lufthansa pour des problèmes ayant jusqu'à 106 vols par semaine et comparent avantageusement leurs résultats à ceux provenant de méthodes existantes.

REXING *et al.* (2000) proposent quant à eux une méthode de résolution pour le problème d'affectation des types d'appareils aux vols sur une journée, sans contrainte supplémentaire, où les modifications à l'horaire sont discrétisées. Les auteurs ont des réserves quant aux modifications de 20 minutes, mais croient que la demande est

indépendante de l'horaire si les changements sont d'au plus une dizaine de minutes. Ils permettent des départs à toutes les minutes ou aux cinq minutes à l'intérieur de l'intervalle choisi. Cette discrétisation peut être spécifique aux segments de vol. Les auteurs ont remarqué que l'augmentation de la taille du réseau avec des départs à toutes les minutes n'était pas justifiée. Les solutions fournies avec des départs aux cinq minutes sont de presque aussi bonne qualité. Les coûts des revenus perdus sont estimés par une courbe normale. Les auteurs utilisent les techniques d'agrégation de HANE *et al.* (1995). De plus, après avoir agrégé le réseau, certains arcs représentant différents horaires se trouvent à partager le même noeud de départ et on peut ne garder que celui qui se termine le plus tôt parmi ceux qui ont le même coût. L'article propose deux méthodes de résolution. La première est la méthode directe de HANE *et al.* (1995) avec des fenêtres de temps. Cette méthode est simple et rapide, mais elle inclut des copies d'arcs pour représenter toutes les différentes possibilités d'horaire. Comme certaines de ces copies sont probablement inutiles, elle utilise beaucoup de mémoire et devient lourde pour les problèmes de grande taille. Les auteurs suggèrent une méthode itérative qui génère les copies d'arcs au besoin. Un problème-maître résout une relaxation du problème alors qu'un sous-problème génère les arcs les plus susceptibles de rendre la solution du problème-maître réalisable pour le problème d'origine. La méthode itérative est très coûteuse en temps de calcul. Par ailleurs, cette méthode fonctionne bien quand les fenêtres de temps sont étroites et qu'elles permettent plusieurs départs successifs. Les auteurs privilégient la méthode directe lorsque les intervalles sont plus larges et que les départs ne sont qu'aux cinq minutes. Les auteurs ont testé leurs méthodes avec des données provenant de United Airlines. Le plus gros problème comprend 2000 vols par jour avec onze types d'appareils et prend quelques heures à résoudre. Les auteurs rapportent que leur méthode permet à la compagnie d'économiser au moins 65 000 dollars américains par jour grâce à la flexibilité accrue. L'heure de départ d'environ seulement 8% des segments est déplacée,

alors que de 10% à 20% des segments ont changé de type d'appareil. Ce sont ces changements de type qui procurent la majeure partie des gains.

Plus récemment, ROSENBERG *et al.* (2004) ont proposé un modèle plus robuste, soit un modèle d'affectation des types d'appareils avec isolation de hub et des cycles courts. La connectivité à un hub est le nombre de segments de vol dans une route d'avions qui font partie d'une sous-route qui débute au hub, se termine à un autre hub, en passant uniquement par des aéroports intermédiaires. Comme certains vols doivent inévitablement être annulés le jour de l'opération, et que ces annulations impliquent habituellement l'annulation d'un cycle complet, soit une séquence de segments de vol qui débutent et se terminent au même aéroport, une affectation de types d'appareils qui facilite la création de rotations d'avions comprenant plusieurs cycles courts ainsi que la réduction de la connectivité aux hubs rendent les solutions moins sensibles aux modifications imprévues. Les auteurs déterminent une borne inférieure sur le nombre de cycles courts en utilisant le réseau de connectivité au hub d'une affectation de types d'appareils. Ils proposent aussi un modèle d'affectation de types d'appareils qui enchâsse plusieurs cycles courts et réduit la connectivité au hub dans une même solution. Les auteurs montrent, grâce à un simulateur très performant, qu'un tel modèle performe mieux, pendant le jour d'opération, que les modèles d'affectation de types d'appareils traditionnels.

Le problème du choix des séquences combinées est étudié par JARRAH et STREHLER (2000). Les auteurs se concentrent sur les séquences combinées qui n'impliquent que deux segments de vol et sur le problème sur une journée dans un réseau de type *hub-and-spoke*. La méthode proposée sélectionne les meilleures séquences, c'est-à-dire celles qui maximisent le revenu espéré parmi les combinaisons de vols possibles. Les estimations de ces revenus sont réalisées à l'aide des prévisions de la demande sans les séquences combinées et des prévisions lorsqu'on ajoute une à une les combinaisons. Pour être réalisable, une combinaison de segments de vol doit respecter un temps

TAB. 2.1 Caractéristiques des principaux modèles d'affectation des types d'appareils

Auteurs	Type de problème	Type de réseau	Fonction-objectif	Contraintes supplém.	Approche de résol.	Taille des probl.
ABARA (1989)	Court courr.	Connexions nb. restreint	Bénéfices Coûts d'opération Déséquil. aux stations	Affectations spéciales Repos de nuit Segments potentiels	Simplexe	2300 vols 150 villes 500 avions
HANE et al. (1995)	Court courr. Quotidien	Lignes de temps Agrégations	Coûts d'opération Revenus perdus	Séquences combinées Connexions courtes	Simplexe Branchement heuristique	2500 vols 11 types 150 villes
CLARKE et al. (1996)	Court courr. Quotidien Considér. entretien équipages	Lignes de temps Agrégations	Coûts d'opération Revenus perdus	Couvre-feu Disponibilité des portes	Simplexe Branchement heuristique	2572 vols 11 types
RUSHMEIER et KONTOGIORGIS (1997)	Quotidien Considér. entretien équipage	Connexions Agrégations	Revenu espéré Coûts d'opération Pénalités non-lin. selon temps de connexion	Séquences combinées Repos de nuit	Simplexe Branchement heuristique	1620 vols 8 types 100 villes
DESAULNIERS et al. (1997)	Moyen courr. Quotidien Fenêtres de temps	Connexions Agrégations	Profits anticipés Pénal. pour connexions courtes	Couvre-feu Séquences combinées Capacité d'entretien Segments potentiels	Gén. de col. Sous-probl. de réseau Branchement heuristique	383 vols 9 types
REXING et al. (2000)	Quotidien Fenêtres de temps	Lignes de temps Agrégations	Coûts d'opération Revenus perdus		Simplexe Génération itér. d'arcs d'horaire	2037 vols 11 types
ROSENBERGER et al. (2004)	Quotidien	Lignes de temps Agrégations	Coûts d'opération des avions Isolation des hubs Cycles courts		Gén. de col. Branchement heuristique	2558 vols 9 types 8 hubs

de connexion minimal et maximal, ne doit pas augmenter le nombre total d'avions nécessaires à la couverture des vols et ne doit pas faire augmenter de façon exagérée le temps de vol par rapport à un vol direct. De plus, les compagnies aériennes désirent équilibrer les allées et les retours (une séquence combinée de Montréal à Vancouver et une autre de Vancouver à Montréal) et restreindre les séquences combinées dans un même marché de façon à s'assurer qu'elles soient suffisamment séparées les unes des autres et des vols directs. Le problème est formulé comme un ensemble de problèmes de réseau liés par des contraintes supplémentaires. Il est résolu comme un problème d'affectation modifié entre les départs et les arrivées aux aéroports principaux. On respecte les contraintes de réalisabilité des combinaisons à l'aide des arcs du réseau (en en omettant certains) ou par des pénalités dans l'objectif. La méthode proposée a été implantée par United Airlines et les bénéfices du modèle sont estimés à environ dix millions par année grâce à la réduction de l'effort nécessaire à l'évaluation des séquences combinées, mais aussi à un meilleur choix de celles-ci.

AHUJA et ORLIN (2002) et AHUJA *et al.* (2004) présentent un algorithme de recherche à l'intérieur d'un voisinage très grand pour résoudre, respectivement, le problème qui combine l'affectation des types d'appareils et le choix des séquences combinées, et une des ses généralisations donnant la possibilité de modifier l'heure de départ des vols à l'intérieur d'une fenêtre de temps. Les auteurs donnent une formulation en nombres entiers qui ne peut être résolue en utilisant les algorithmes exacts commerciaux. Ils proposent donc un algorithme de recherche heuristique qui permute les types d'appareils entre deux vols affectés à un type différent.

Les caractéristiques des principaux modèles d'affectation des types d'appareils aux vols sont résumées dans le tableau 2.1.

2.2 La création des itinéraires d'avions

Le problème du routage des avions doit respecter les contraintes de couverture des vols et les contraintes d'entretien des avions. Tous les modèles de routage d'avions respectent les entretiens réguliers, mais les entretiens cycliques ne font pas toujours partie de la formulation.

On observe deux manières principales de modéliser les contraintes d'entretien régulier. Comme les vols locaux ont principalement lieu durant la journée, les auteurs qui traitent ce type d'horaire supposent souvent que l'entretien aura naturellement lieu pendant la nuit et génèrent des *journées de travail* pour les avions, c'est-à-dire des séquences de segments de vol. Les journées de travail sont ensuite reliées entre elles de façon à pouvoir insérer les périodes d'entretien pendant les arrêts de nuit. On peut aussi permettre l'entretien pendant le jour. Cette deuxième méthode, requise pour les vols internationaux, augmente la taille du réseau pour les vols locaux mais peut améliorer la qualité des solutions.

Certaines compagnies aériennes préfèrent que tous les avions d'une flotte donnée volent sur la même route pour uniformiser leur usure. Par exemple, pour un problème sur une journée avec 30 avions, une route de 30 jours comprenant chaque segment de vol quotidien est valide si on positionne un avion différent à chaque début de journée. De plus, si l'origine est une station d'entretien cyclique, chaque avion passera nécessairement par cette station au moins une fois à tous les 30 jours. Certains auteurs exploitent la propriété eulérienne du graphe pour imposer l'entretien cyclique.

Le problème de la création des itinéraires d'avions peut être formulé comme un problème de réalisabilité seulement, car les coûts d'opération sont fixes par type d'appareil et il n'y a pas de coût de connexion pour les avions. Certains auteurs

incluent néanmoins des coûts d'entretien car l'objectif de seulement créer des occasions d'entretien peut devenir inutilement coûteux. Par exemple, si les délais maximaux pour l'entretien sont de quatre jours et que les occasions d'entretien sont aux trois jours, on ne peut pas passer outre aucune séance prévue et au bout de douze jours, une séance superflue a été réalisée.

RICHARDSON (1976) étudie un problème particulier de routage d'avions et d'affectation des passagers aux itinéraires qui s'adresse aux compagnies aériennes avec des vols internationaux dans un marché à faible densité. L'auteur propose une méthode de décomposition de Benders pour résoudre le problème. Lorsque les variables binaires de connexion sont fixées temporairement pour un itinéraire, le problème se réduit à un programme linéaire standard qui affecte les passagers aux itinéraires. La méthode a été testée avec des données de Qantas Airlines pour 26 villes de la région du Pacifique.

DASKIN et PANAYOTOPOULOS (1989) s'attaquent au problème sur une journée avec un aéroport principal unique. Les itinéraires ne comportent généralement la visite que d'un seul aéroport secondaire avant de revenir à l'aéroport principal. De plus, on connaît l'heure de retour, le temps d'entretien requis et le profit de chaque itinéraire. Des tranches de temps sont définies afin de déterminer les itinéraires incompatibles, c'est-à-dire ceux qui ne peuvent pas être affectés au même avion. Les auteurs utilisent une relaxation lagrangienne où les contraintes d'incompatibilité sont dualisées. Le sous-problème peut être résolu par inspection et on utilise une méthode de sous-gradient pour le dual lagrangien. Comme la solution de la relaxation lagrangienne viole souvent les contraintes d'incompatibilité, les auteurs suggèrent des heuristiques d'échanges pour la rendre réalisable. Ils résolvent des problèmes générés aléatoirement d'au plus 100 itinéraires pour lesquels la méthode proposée donne de bons résultats quand le nombre d'avions requis est approximativement égal au nombre d'avions disponibles. Ils résolvent donc des modèles simplifiés et de petite taille.

FEO et BARD (1989) étudient le problème du routage des avions et de la localisation des stations d'entretien avec comme objectif de trouver le nombre minimal de stations d'entretien requises pour respecter les séances régulières aux quatre jours. Les auteurs supposent que les caractéristiques de chaque station sont identiques. Ils forcent les séances d'entretien à avoir lieu la nuit et ils utilisent la notion de journées de travail, qui doivent être fournies comme données d'entrée. Ils formulent le problème comme un modèle de réseau multi-flots en nombres entiers à coût minimal. Chaque avion représente un type de flot différent et les journées de travail sont sur les noeuds du réseau. Les auteurs proposent deux approches. La première approche résout le problème avec un horizon d'une semaine en cherchant un chemin contenant sept arcs de coût total minimal et respectant des conditions initiales et finales. Pour introduire les contraintes d'entretien, chaque noeud est dédoublé et deux arcs relient les deux copies, un pour l'entretien avec une capacité finie et un autre avec capacité infinie. Une contrainte impose un maximum de quatre jours entre chaque visite. Une méthode heuristique (*FIFO*) est utilisée pour générer un ensemble de journées de travail. Ces journées de travail ne respectent pas nécessairement les contraintes d'entretien. On résout en augmentant d'une journée l'horizon à chaque itération. Lorsque la fin de l'horizon est atteinte, on diminue d'une unité le nombre de stations d'entretien et on recommence. Une seconde approche est aussi proposée. Il s'agit d'une heuristique en deux phases sur un horizon infini qui génère premièrement des cycles couvrant les vols en exploitant la propriété eulérienne du graphe et localise ensuite les stations d'entretien à l'aide d'échanges par paire. Les auteurs ont testé leurs méthodes à l'aide des données d'American Airlines pour la flotte de Boeing 727. Les données comportaient 154 journées de travail quotidiennes reliant 76 villes pendant la semaine. La méthode a permis de diminuer le nombre de stations d'entretien de 22 à 17.

GOPALAN et TALLURI (1998) s'intéressent au problème de la création des itinéraires d'avions où chaque avion doit avoir un entretien régulier aux trois jours et un entretien

cyclique au moins une fois pendant la période de planification. Les auteurs proposent un algorithme qui fonctionne en temps polynômial. Ils considèrent le problème des vols locaux sur une journée avec un modèle ayant un horizon infini et le problème des vols internationaux à l'aide un modèle ayant un horizon égal au nombre d'avions de la flotte. Leur méthode cherche des cycles (ou des chemins) qui débutent et se terminent à (ou passent par) la station d'entretien cyclique. Pour débiter la procédure, on doit avoir en mains un ensemble de journées de travail. Chacune de ces séquences de vol est construite en fixant des connexions à partir de règles simples comme *FIFO* ou fournies par les planificateurs. Le réseau contient donc seulement des stations où il existe un repos de nuit et les itinéraires sont formés sur ce réseau simplifié en respectant les entretiens réguliers aux trois jours. Le fait d'avoir fixé certaines connexions pour générer les journées de travail avant de construire les itinéraires peut créer des *itinéraires fermés*, qui ne passent jamais par une station d'entretien cyclique. La méthode proposée tente d'*ouvrir* les itinéraires par des échanges de segments ou même de types d'appareils à l'aide des méthodes proposées par TALLURI (1996). Pour respecter les séances d'entretien régulier, on transforme le réseau de façon à ce qu'il soit impossible qu'un chemin passe par trois repos de nuit de suite dans des stations qui ne font pas d'entretien régulier. Cette transformation rend le problème polynômial, car il ne demande ensuite qu'une recherche de cycles eulériens dans le réseau modifié. L'astuce de modélisation proposée par les auteurs ne fonctionne, par contre, que pour les problèmes où l'entretien doit avoir lieu à tous les trois jours maximum. Les auteurs montrent que si le délais entre les séances peut être de quatre jours ou plus alors le problème de la création des itinéraires d'avions est NP-complet.

TALLURI (1998) étudie le problème du routage d'avions lorsque l'entretien régulier doit avoir lieu au plus à tous les quatre jours. Cet article poursuit le travail présenté précédemment par GOPALAN et TALLURI (1998). L'auteur montre que s'il n'y a qu'une station d'entretien régulier, alors le problème peut être résolu par un

algorithme de couplage en un temps polynômial. Par ailleurs, il montre que lorsque l'entretien cyclique n'est pas requis, le problème peut aussi être résolu en un temps polynômial peu importe le nombre de stations d'entretien. L'auteur suggère ensuite deux heuristiques pour résoudre le problème général. La première méthode proposée partitionne le graphe original en un ensemble de cycles qui a) comprennent chacun un cycle eulérien, b) respectent l'entretien régulier aux quatre jours et c) peuvent partager des noeuds. Elle essaie ensuite d'obtenir un seul cycle eulérien en les reliant entre eux. La deuxième méthode trouve un cycle eulérien et tente de briser les séquences qui violent les contraintes d'entretien régulier.

Dans l'article de CLARKE *et al.* (1997), le problème du routage d'avions est formulé comme un problème de voyageur de commerce asymétrique avec des contraintes supplémentaires pour l'entretien régulier. Ces contraintes sont en fait des coupes servant à empêcher les chemins qui ne respectent pas ce type d'entretien. Le problème est résolu par une relaxation lagrangienne et une méthode de sous-gradient. Les auteurs modélisent les bénéfices générés par le choix de séquences combinées. On affecte à ces séquences un coût qui dépend du temps passé au sol entre les vols et la différence entre le temps total du voyage et le temps d'un vol direct pour la même destination finale. L'entretien cyclique est assuré lorsque l'itinéraire formé débute et se termine à la station d'entretien. Les contraintes d'élimination de sous-tours du problème de voyageur de commerce forcent l'itinéraire à avoir cette propriété. Un itinéraire est réalisable s'il n'est pas *brisé* et si la durée la plus longue entre deux entretiens réguliers est inférieure à un intervalle de temps spécifié. La méthode utilise certaines agrégations de noeuds et d'arcs pour simplifier la structure du réseau en ne gardant qu'un super-noeud par ville. La relaxation lagrangienne dualise les contraintes d'élimination de sous-tours et les contraintes supplémentaires d'entretien régulier, mais seulement une à la fois, lorsqu'elles sont identifiées. La meilleure solution réalisable pour le problème d'origine est gardée en mémoire jusqu'à ce que le dual

lagrangien soit résolu. La solution finale est donc réalisable, mais pas nécessairement optimale. Les auteurs ont testé leur méthode avec des données provenant de onze flottes différentes de la compagnie Delta Air Lines comprenant approximativement de 40 à 1000 noeuds et de 40 à 4000 arcs. L'approche suggérée semble avoir bien fonctionné pour certains problèmes dont la solution a par la suite été prouvée optimale par une méthode de séparation et d'évaluation, mais ne résout pas efficacement les problèmes les plus difficiles.

Un cas particulier du problème de routage d'avions est le routage de jets en temps partagé. Certains fabricants vendent leurs avions à plusieurs clients différents qui se partagent les appareils. Les clients sont donc des propriétaires partiels d'un type d'appareil avec un nombre fixe d'heures de vol *achetées* par année. Ces fabricants doivent gérer l'utilisation des appareils. Certaines requêtes, qui consistent en une origine, une destination et un moment de départ, sont connues à l'avance alors qu'un ensemble de nouvelles requêtes peut arriver à tout moment. À un moment donné, les avions sont à différents endroits dans le réseau (au sol ou en vol) et l'objectif est de satisfaire toutes les requêtes des clients, dont le nombre d'heures d'utilisation n'est pas épuisé, à un coût minimal, en respectant des contraintes d'entretien pour les appareils. Les coûts comprennent les coûts opérationnels, les vols de repositionnement à vide de même que le coût de location d'appareils et de personnel supplémentaires, s'il le faut. On doit donc décider de l'affectation des requêtes aux avions et créer les itinéraires des avions. Ce problème est traité par KESKINCAK et TAYUR (1998). Les auteurs modélisent les contraintes d'entretien à l'aide d'un nombre d'heures de vol et atterrissages à ne pas dépasser ainsi que des séances d'entretien fixées comme des requêtes. Un modèle linéaire en nombres entiers résolu par l'algorithme du Simplexe est efficace pour les petits problèmes comprenant jusqu'à 20 avions et 50 requêtes. Pour les plus gros problèmes comprenant jusqu'à 30 avions et 80 requêtes, les auteurs proposent une heuristique qui crée des horaires partiels et qui les combine et les

améliore avec des échanges. Cette heuristique sépare les avions en trois groupes, ceux qui n'auront pas besoin d'entretien pendant la période, ceux dont l'entretien est fixée comme une requête et les autres avions. La méthode proposée a été testée avec des données générées aléatoirement selon un modèle réel et les solutions, qui ne comprennent jamais de location d'avions ou de personnel que la solution optimale n'inclut pas, se trouvent en moyenne à 2% de l'optimum.

Les caractéristiques des principaux modèles de routage d'avions sont résumées dans le tableau 2.2.

TAB. 2.2 Caractéristiques des principaux modèles de routage d'avions

Auteurs	Type de problème	Modèle	Fonction objectif	Contraintes supplém.	Approche de résolution
DASKIN et PANAYOTOPOULOS (1989)	Quotidien Un seul hub Entretien entre les itinér.	Tranches de temps Incompatibilités	Profits itinéraire/avion		Relaxation lagrangienne Heuristique
FEO et BARD (1989)	Localisation des stations d'entretien Entretien de nuit	Partitionnement d'ensemble	Coûts fixes et coûts variables d'entretien	Entretien rég : 4 jours cyclique	Cycles eulériens Heuristique d'échanges
CLARKE et al. (1997)	Quotidien Entretien de jour ou de nuit	PVC asymétrique Agrégations	Profits des séquences combinées	Entretien rég : 3-4 jours cyclique	Relaxation lagrangienne Contraintes dualisées au besoin
GOPALAN et TALLURI (1998)	Quotidien Périodique Entretien de nuit	Tranches d'une journée	Aucun coût	Entretien rég : 3 jours cyclique	Polynômial Cycles eulériens
TALLURI (1998)	Quotidien Périodique Entretien de nuit	Tranches d'une journée	Aucun coût	Entretien rég : 4 jours cyclique	Cycles eulériens Heuristique d'échanges

2.3 La création des rotations d'équipages

La plupart des approches proposées pour la création des rotations d'équipages sont basées sur la résolution d'un problème de partitionnement d'ensemble dans lequel les

variables de décision représentent les rotations possibles. Pour pouvoir utiliser cette formulation, les rotations doivent pouvoir être énumérées ou générées dynamiquement. Énumérer toutes les rotations peut devenir une tâche fastidieuse étant donné le très grand nombre de rotations possibles et le nombre élevé de règles devant être vérifiées pour assurer leur validité.

L'énumération des rotations est parfois faite en énumérant au préalable les journées de travail possibles qui correspondent à l'horaire et en combinant ensuite ces journées de travail pour former les rotations. L'énumération des journées de travail peut être accomplie avec un algorithme de recherche. Par exemple, par une recherche en profondeur d'abord, un arbre est construit pour chacun des segments de vol. La racine d'un arbre représente le segment correspondant et il y a un descendant par connexion possible qui a lui-même un descendant par connexion subséquente possible. La profondeur de l'arbre de recherche est limitée par le nombre maximal de segments de vol permis dans une journée par les règles de travail. Chaque chemin de la racine à un descendant peut représenter une journée de travail s'il ne viole aucune règle de validité du contrat de travail. On pourrait de la même façon énumérer toutes les rotations en mettant les journées de travail sur les noeuds d'un arbre, mais la grande taille d'un tel arbre rend cette recherche impraticable. À cause du nombre élevé de variables dans le problème de partitionnement d'ensemble des problèmes réels, des approches d'optimisation locale ont souvent été utilisées.

ANBIL *et al.* (1992) s'intéressent à un problème de rotations d'équipages chez American Airlines qui comprend quotidiennement 800 segments de vol domestiques et probablement des milliards de rotations possibles. Dans le but d'obtenir une meilleure estimation de l'optimum global, les auteurs trouvent une solution optimale de la relaxation linéaire du problème comprenant un grand sous-ensemble de rotations. Parmi les douze millions de rotations générées, environ cinq millions et demi sont gardées après avoir éliminé les doubles et les rotations de coût très élevé. Leur méthode

de résolution ajoute ensuite par milliers les rotations au problème de partitionnement d'ensemble et ne conserve, à chaque itération, que les variables de base jusqu'à ce que toutes les rotations du sous-ensemble aient été considérées. La méthode évalue ensuite le coût réduit des rotations exclues du sous-ensemble. Les auteurs croient que leur méthode fournit une bonne solution à la relaxation linéaire du problème d'origine, sans pour autant pouvoir en garantir la qualité. Ils doivent ensuite imbriquer cette méthode à l'intérieur d'une approche de séparation et d'évaluation pour trouver une solution entière. Le branchement se fait sur les connexions en fixant deux segments de vol qui se suivront. Des rotations du sous-ensemble peuvent être ajoutées à chaque noeud de branchement si le nombre de variables diminue beaucoup ou si le problème devient non réalisable. La méthode proposée utilise une fonction de coût sophistiquée comprenant premièrement la différence entre le nombre minimal d'heures garanties et la durée réelle de vol, deuxièmement, le coût des connexions longues entre les segments et les journées de travail et, finalement, le coût des vols de repositionnement. Les contraintes sont celles retrouvées habituellement dans le contrat de travail. Le modèle comprend aussi des contraintes d'équilibrage du temps de vol entre les bases et des contraintes qui limitent le nombre de changements d'appareil. La méthode a permis à la compagnie aérienne de diminuer de 5 à 11% les coûts en excès.

GRAVES *et al.* (1993) décrivent le problème des rotations d'équipages chez United Airlines. Pour trouver une solution initiale, le système utilise une variation de la formulation de partitionnement d'ensemble qui permet aux segments de vol de ne pas être couverts ou de l'être à plusieurs reprises, en mettant toutefois une pénalité dans la fonction-objectif. Une fois une solution initiale trouvée de cette façon, une optimisation locale tente de l'améliorer. Un générateur de rotations fait une énumération intelligente des possibilités et un optimiseur sépare le problème en plusieurs sous-problèmes qui sont résolus individuellement. Des rotations disjointes sont ainsi obtenues que la méthode améliore à l'aide d'échanges simples. Le système

semble résoudre efficacement des problèmes qui comprennent jusqu'à 1700 segments de vol quotidiennement. Des économies annuelles d'environ 16 millions de dollars américains ont été réalisées suite à l'implantation du système chez United Airlines.

HOFFMAN et PADBERG (1993) trouvent des solutions optimales à plusieurs problèmes fournis par des compagnies aériennes de grande taille en utilisant un algorithme de séparation et de coupes (*branch-and-cut*). Les problèmes comprennent un sous-ensemble de rotations générées par des heuristiques favorisant les rotations de plus petit coût. Leur algorithme comprend quatre composantes : a) un préprocesseur qui localise les contraintes dominées, b) une heuristique qui fournit des solutions réalisables entières rapidement en décomposant le problème, c) un générateur de coupes qui trouve des inégalités valides à chaque noeud de l'arbre de branchement en identifiant certaines cliques sur un graphe, et d) une méthode de branchement qui utilise la densité de la matrice, le nombre de variables fractionnaires par contrainte, les variables dont la valeur s'approche de 1 et les variables qui modifieraient de beaucoup la valeur de l'objectif. Les auteurs ne spécifient pas les détails des règles de validité qu'ils incorporent à leur modèle, mais affirment que leur approche peut considérer explicitement des contraintes globales pour les bases des équipages.

Les approches les plus récentes peuvent résoudre les problèmes de rotations d'équipages à l'optimalité en ne générant que les rotations les plus intéressantes. Le nombre de rotations générées est souvent très petit par rapport au nombre total de rotations possibles. Comme le problème est en nombres entiers, on doit inclure l'approche de *génération de colonnes* dans une méthode de séparation et d'évaluation, ce qui rend la résolution plus ardue. Les auteurs développent des méthodes de branchements heuristiques pour accélérer la résolution.

La première méthode de génération de colonnes pour le problème de la création des rotations d'équipages a été proposée par LAVOIE *et al.* (1988). Il a ensuite fallu attendre

quelques années pour que d'autres auteurs reprennent ce type d'approche. Cette première méthode, développée pour Air France, requiert premièrement l'énumération complète des journées de travail. Un réseau est ensuite créé en mettant les journées de travail sur les noeuds. Plusieurs copies du même noeud peuvent être créées et des arcs de repos les relient de façon à ce que chaque chemin dans ce réseau augmenté représente une rotation valide (les journées de travail possibles suite à une journée de travail donnée dépendent des journées précédentes). On modélise ainsi le repos de nuit minimal, le temps total hors de la base maximal et le nombre d'heures de vol maximal. Leur approche résout un premier problème de partitionnement d'ensemble à l'aide d'une solution initiale fournie par la compagnie. Les variables duales sont ensuite utilisées dans un sous-problème de plus court chemin dans un réseau afin de générer des rotations de coût réduit négatif. La méthode itère ensuite entre le problème-maître de partitionnement d'ensemble et le sous-problème de réseau jusqu'à ce qu'aucune autre rotation ne puisse être générée. Une fonction de coût simplifiée comprenant le temps total hors de la base et les repos de nuit est utilisée. Les problèmes de long et de moyen courrier pour les agents de bord d'Air France traités avec la méthode comprenaient jusqu'à 329 segments de vol quotidiens et les solutions ont permis des économies moyennes d'environ 4 à 5% par rapport aux solutions manuelles de la compagnie. Par ailleurs, les solutions étant, sauf exceptions, entières naturellement, les auteurs n'ont pas eu à développer de méthode de séparation et d'évaluation.

BARNHART *et al.* (1995) étudient le problème du choix des vols de repositionnement pour les rotations d'équipages dans un cadre de vols de longue durée. Le réseau des vols internationaux étant plus épars, il se crée des périodes d'inactivité prolongée pour les équipages qui doivent prendre leurs repos à des aéroports où le trafic est plus faible. Il est alors critique de bien choisir les vols de repositionnement qui deviennent souvent avantageux dans cette situation. Si l'on choisit les vols de repositionnement parmi tous les vols de l'ensemble des compagnies aériennes, on s'assure de trouver

la meilleure solution possible, mais ces millions de possibilités sont très difficiles à gérer. Les auteurs proposent donc une méthode heuristique d'élimination successive qui sélectionne les vols de repositionnement potentiels en évaluant le bénéfice associé à chacun des vols. L'approche développée par les auteurs résout premièrement la relaxation linéaire d'un problème de rotations d'équipages qui respecte les règles de validité mais qui ne comprend pas de contrainte sur la disponibilité des équipages aux bases. La résolution se fait par une méthode de génération de colonnes dont la fonction de coût ne comprend que le temps total hors de la base et les vols de repositionnement. La méthode sélectionne ensuite des vols de repositionnement à l'aide des coûts réduits de rotations partielles (de la base de l'équipage à la station de départ de certains vols de repositionnement ou de la station d'arrivée des vols de repositionnement à la base de l'équipage) et les ajoute au problème des rotations. Les vols de repositionnement potentiels sont évalués à l'aide d'un calcul très simple, ce qui permet l'évaluation rapide d'un grand nombre d'entre eux. La méthode itère ainsi jusqu'à ce que la valeur de la solution de la relaxation linéaire soit stable. La méthode a été testée avec des données provenant d'American Airlines comprenant jusqu'à 833 segments de vol avec deux bases parmi 41 aéroports sur un horizon de neuf semaines. Le temps maximal hors de la base pour un équipage est fixé à 15 jours. Les auteurs rapportent des solutions qui diminuent les crédits de vol par rapport aux solutions fournies par la compagnie aérienne pour environ cinq millions de dollars américains. Ces économies sont en partie attribuables à une diminution du nombre d'heures des vols de repositionnement utilisées.

HOUSOS et ELMROTH (1997) présentent un modèle, employé entre autres par la compagnie aérienne Lufthansa, qui utilise les connaissances des planificateurs pour fixer les moments et les endroits où un équipage change d'avion ou pour fixer des connexions. Les auteurs soulignent l'importance de générer des solutions acceptables pour les planificateurs et d'obtenir rapidement des solutions car les départements de

planification aiment retarder le plus possible la résolution du problème des rotations pour que l'horaire représente mieux la demande. La méthode proposée sélectionne donc des sous-problèmes qui réduisent la taille du problème en générant toutes les rotations d'une journée et en fixant les décisions des journées qui précèdent. Ces sous-problèmes distincts sont ensuite optimisés séparément. Les auteurs pensent obtenir des solutions proches de l'optimum global, bien que la tendance soit plutôt à optimiser globalement en mettant des efforts pour développer de bonnes méthodes de résolution.

CHU *et al.* (1997) posent le problème des rotations d'équipages comme un problème de partitionnement d'ensemble où chaque colonne représente un cycle dirigé dans un réseau qui débute et se termine à une base. La méthode proposée comporte trois étapes. Premièrement, une série de petits programmes linéaires sont résolus successivement et de nouvelles colonnes sont générées à l'aide d'un graphe jusqu'à l'obtention d'environ vingt millions de colonnes. Une solution optimale de la relaxation linéaire du problème comprenant ces vingt millions de colonnes est ensuite trouvée et, finalement, un branchement heuristique est appliqué aux 15 000 des meilleures colonnes. Les variables de branchement sont celles associées aux arcs de connexion et on fixe de préférence les connexions aux aéroports secondaires et aux segments de vol qui ont peu de prédécesseurs ou de successeurs possibles. Les auteurs rapportent des solutions dont l'écart d'intégrité est de moins de 2% pour des données provenant des avions de type MD-80, la plus grosse flotte d'American Airlines. Chez American Airlines, les crédits de vol ne sont plus que de 0,2% avec ces nouvelles améliorations au système de planification des rotations d'équipages.

DESAULNIERS *et al.* (1998a) mettent en évidence la structure commune de plusieurs problèmes d'horaires d'équipages et proposent une méthodologie générale de résolution pour obtenir des solutions proches de l'optimum ou optimales dans certains cas. Les problèmes de rotations d'équipages font partie de ces problèmes d'horaires, de même que les problèmes d'affectation des rotations au personnel pour la création d'ho-

raires mensuels et les modifications opérationnelles. La structure commune comprend un réseau de type espace-temps, où les chemins dans le réseau correspondent à des horaires réalisables. On tient compte des contraintes qui restreignent la réalisabilité des chemins soit dans la conception du réseau, par exemple pour le temps de connexion minimal, soit pendant la construction des chemins via l'utilisation de variables de ressources, par exemple pour le temps total hors de la base maximal ou le nombre atterrissages maximal. Les contraintes globales qui assurent la couverture des tâches ou une distribution équilibrée des rotations entre les bases brisent la structure de réseau. Les auteurs proposent une méthode de génération de colonnes imbriquée dans une méthode de séparation et d'évaluation pour résoudre ces problèmes d'horaires. À chaque noeud de branchement, on résout la relaxation linéaire du problème afin d'obtenir une borne inférieure. On résout alternativement un problème-maître restreint (qui inclut un problème de partitionnement d'ensemble et des contraintes globales) et plusieurs sous-problèmes (un par équipage). Le problème-maître restreint, résolu par la méthode du Simplexe, est mis à jour à chaque itération et vise à trouver la meilleure solution en tenant compte des chemins disponibles. Chaque sous-problème est un problème de plus court chemin avec des variables de ressources qui permettent la génération de tous les chemins réalisables. Le rôle d'un sous-problème est de générer des chemins réalisables intéressants (de coût réduit négatif) en utilisant l'information duale associée à la solution du problème-maître. Les sous-problèmes sont résolus par un algorithme de programmation dynamique. Les noeuds qui forment le réseau représentent des segments de vol ou des journées de travail. Les auteurs remarquent que l'utilisation des noeuds représentant les segments est bonne quand le nombre de journées de travail est très grand (par exemple pour des transporteurs régionaux) car elle évite l'énumération explicite des journées, mais requiert l'utilisation d'un plus grand nombre de variables de ressources. La fonction de coût (séparable par équipage) et les fonctions d'évaluation de la valeur des ressources peuvent être non linéaires, non convexes ou discontinues tant qu'elles demeurent non décroissantes. Les

auteurs proposent différentes stratégies de branchement optimales, comme imposer que deux tâches soient exécutées consécutivement par le même équipage, ou heuristiques, comme fixer des chemins sans inclure la branche qui l'interdit. Ils proposent aussi de prendre plusieurs décisions à la fois ou de limiter l'ensemble des états explorés par l'algorithme de programmation dynamique pour accélérer la méthode de branchement.

DESAULNIERS *et al.* (1997a) présentent une application de la méthode unifiée décrite par DESAULNIERS *et al.* (1998a) chez Air France. Le problème est celui des moyens courriers sur une semaine avec des conditions initiales et finales pour les équipages au début et à la fin d'une semaine. En plus des contraintes normales sur les journées de travail et les rotations et des contraintes globales aux bases, le modèle général permet d'ajouter certaines variables supplémentaires auxquelles on associe des pénalités dans la fonction-objectif. Une de ces variables supplémentaires pourrait, par exemple, servir à imposer une limite sur le nombre de changements d'avion. Des copies des journées de travail sont créées, comme l'avaient fait LAVOIE *et al.* (1988), pour inclure des contraintes qui dépendent de l'état de l'équipage à un moment dans le temps. Les différents états possibles permettent, par exemple, un repos de nuit plus court si le précédent était très long. Les auteurs présentent des manières de réduire la taille du réseau par des agrégations de noeuds et par l'élimination récursive des noeuds sans prédécesseur ou successeur. Les vols de repositionnement sont possibles, mais seul les vols de la compagnie elle-même et certains autres vols spécifiques sont inclus. Les onze problèmes testés comprenaient de 154 à 1157 segments de vol et la méthode proposée a permis de réduire en moyenne de 6,24% les coûts d'équipages.

Au lieu de considérer la formulation traditionnelle qui choisit un ensemble de rotations qui partitionnent les segments de vol, VANCE *et al.* (1997) présentent un modèle qui sépare en deux étapes le processus de décision. La première étape sélectionne un ensemble de journées de travail qui couvrent tous les segments de vol et la deuxième, un

ensemble de rotations qui partitionnent les journées de travail. Les auteurs présentent une nouvelle méthode de décomposition à partir d'une formulation contenant à la fois des variables représentant les journées de travail et des variables représentant les rotations. Cette méthode de décomposition fournit une borne plus serrée et respecte la propriété d'intégrité plus souvent. Cette approche augmente la taille du problème à résoudre mais rend la matrice des coefficients moins dense. La fonction de coût ne comprend que les coûts en excès. Les auteurs ont testé leur méthode avec des données provenant d'American Airlines pour deux petits problèmes qui comprenaient 144 et 174 segments de vol.

BARNHART et SHENOI (1998) présentent une approche qui résout premièrement un modèle approximatif du problème des rotations d'équipages pour le long courrier, sur une semaine, et qui utilise ensuite cette solution comme point de départ à une méthode conventionnelle de génération de colonnes. Les auteurs utilisent une structure de coût simplifiée qui comprend le temps total hors de la base et le coût des vols de repositionnement. Le modèle approximatif ne tient pas compte des contraintes qui restreignent le temps total hors de la base et les possibilités pour les stations de début et de fin des rotations. Sans ces contraintes, les variables représentant les rotations peuvent être remplacées par des variables représentant les journées de travail, ce qui permet l'utilisation d'un réseau à lignes de temps dans lequel certaines agrégations de noeuds, comme celles décrites par HANE *et al.* (1995), peuvent être faites. Ces relaxations diminuent beaucoup la taille du problème, mais ne donnent pas nécessairement de solution réalisable pour le problème d'origine. Ensuite, les auteurs présentent une méthode de génération de colonnes qui utilise les colonnes optimales de la solution approximative et éventuellement toutes les colonnes. Les tests faits à l'aide de données provenant de problèmes d'une grande compagnie américaine ayant jusqu'à 875 vols par semaine ont montré que le modèle donnait une borne inférieure rapidement et souvent proche de l'optimum (1%). Réduire ainsi

le temps de résolution permet à la méthode de passer plus de temps à choisir de bons vols de repositionnement. Les auteurs utilisent pour cela le générateur de vols de repositionnement présenté par BARNHART *et al.* (1995). Le modèle n'inclut pas de contraintes de disponibilité des équipages aux bases.

Certains auteurs ont étudié le problème d'accroître le niveau de robustesse des rotations d'équipages. Compte tenu de la forte densité du trafic aérien et de l'imprévisibilité du climat, certains vols doivent inévitablement être retardés, ou même annulés, le jour de l'opération. Ceci occasionne des coûts additionnels pour les compagnies aériennes et bien des inconvénients pour les passagers. Les rotations d'équipages habituellement construites ont tendance à provoquer des retards en chaîne pour les vols qui suivent la perturbation imprévue. Une rotation d'équipages est dite plus robuste qu'une autre si elle réduit cet effet domino. EHRGOTT et RYAN (2002) proposent une approche bi-critère qui minimise les coûts d'équipages tout en maximisant la robustesse. En séparant les deux objectifs conflictuels, les auteurs évitent de devoir mettre un prix à la robustesse. Au lieu, ils mesurent la robustesse par le temps au sol (*ground time*). Leurs résultats montrent des gains significatifs en robustesse avec seulement de légères augmentations au niveau des coûts d'équipages. CHEBALOV et KLABJAN (2006) présentent quant à eux un modèle qui produit des rotations d'équipages plus robustes en incluant dans la fonction-objectif un terme qui maximise le nombre d'équipages qui peuvent potentiellement être permutés pendant le jour d'opération.

Les caractéristiques des principaux modèles de rotations d'équipages sont résumées dans le tableau 2.3.

TAB. 2.3 Caractéristiques des principaux modèles de rotations d'équipages

Auteurs	Type de problème	Type de réseau	Fonction objectif	Contraintes supplém.	Approche de résol.	Taille des probl.
ANBIL et al. (1992)	Quotidien	Segments de vol	Min gar./j Conn.longues Repos longs Deadheads Pénalités rotations trop courtes	Vol max :8h/j Max 12h/j Repos de nuit 8 hres en 24 Equil.bases Changements d'avion max	Heuristique itérative Sous-ens. de rotations	800 vols
GRAVES et al. (1993)	Quotidien Court courr.	Segments de vol	Crédits de vol	Conn. min	Heuristique de décompos. Échanges	1700 vols
HOFFMAN et PADBERG (1993)		Segments de vol	Deadheads Per diem des dépenses Temps total hors base	Vol max Hors base max Contraintes aux bases	Séparation et coupes	800 vols 3 bases
LAVOIE et al. (1988)	Moyen et long courr.	Journées de travail	Temps total hors base Repos de nuit	Repos de nuit 8 hres en 24 Hors base max Vol max Deadheads	Génération de colonnes Sous-probl : PCC simple	329 vols
CHU et al. (1997)	Court courr. Quotidien	Segments de vol	Crédits de vol	Conn. min Hors base max max 4j/rot.	Heuristique itérative Sous-ens. de rotations	1000 vols
DESAULNIERS et al. (1997)	Moyen courr. Hebdomad.	Journées de travail Agrégations	Non-lin. Deadheads Pénalités changements d'avion	Conn. min Hors base max Vol max Atterris.max Équil.bases Repos réduit max	Gén. de col. Sous-probl : PCC avec ressources Branchement heuristique	1157 vols 2 bases 63 stations
VANCE et al. (1997)	Court courr. Quotidien	Journées de travail	Coût en excès rotation vs somme des journées	Conn.min :30m Conn.max :4h Max 12h/j Vol max :8h/j Repos de nuit 3 jours/rot.	Décomp. journées / vols	174 vols
BARNHART et SHENOI (1998)	Long courr. Hebdomad.	Journées de travail Agrégations	Temps total hors base Deadheads	Hors base max Repos de nuit Journée de travail max	Gén. de col. approxim. et réelles Branchement heuristique	875 vols 2 bases

2.4 La planification tactique intégrée

Comme nous l'avons expliqué précédemment, utiliser une approche séquentielle réduit considérablement la complexité de la planification, mais peut donner des solutions qui sont très loin de l'optimum. De plus, la perte de flexibilité imposée par certaines décisions prises à une étape donnée peut rendre non réalisables les problèmes des niveaux subséquents. Par exemple, lorsque vient le temps de créer les itinéraires des avions, il est possible de ne pas en trouver qui respecte les contraintes d'entretien si les types d'appareils associés aux segments de vol sont déjà fixés. Bien que ce problème ne survienne que rarement dans le cas des vols locaux car les occasions d'entretien sont nombreuses, créer ainsi des itinéraires d'avions dans un contexte de vols internationaux peut être problématique.

2.4.1 Intégration partielle

Dans la section sur l'affectation des types d'appareils aux segments de vol, nous avons présenté le travail de CLARKE *et al.* (1996), SUBRAMANIAN *et al.* (1994) et RUSHMEIER et KONTOGIORGIS (1997), des auteurs qui ont fait un effort dans la direction de l'intégration en incluant des considérations pour l'entretien et les équipages lors de l'affectation des types d'appareils. Ces auteurs n'incluent aucune contrainte liant les problèmes, mais favorisent les opportunités de sorte que la création des itinéraires d'avions et celle des rotations d'équipages soient facilitées. Les travaux présentés dans la présente section, bien qu'ils ne soient pas totalement intégrés dans la plupart des cas non plus, analysent tous l'impact de l'intégration par rapport à une méthode séquentielle.

BARNHART *et al.* (1998a) tentent l'intégration partielle du choix des types d'avions et de la création des itinéraires d'avions pour des vols de longue durée. Les auteurs utilisent la notion de journée de travail pour les avions, soit une séquence de vols d'une base d'entretien à une autre qui satisfait les contraintes d'entretien. Une *journée de travail augmentée* se définit comme une journée de travail à laquelle on ajoute un temps minimal d'entretien après le dernier vol. L'objectif de la méthode proposée par les auteurs est de sélectionner un ensemble de journées de travail augmentées au moindre coût de façon à a) couvrir chaque segment de vol avec exactement un type d'appareil, b) former un ensemble d'itinéraires d'avions qui débutent et se terminent à la même station d'entretien pour chaque type d'appareil, et c) respecter la taille des flottes. Le coût d'une journée de travail, par appareil, comprend les coûts opérationnels et d'entretien et les revenus (ventes et séquences combinées). Le nombre de journées de travail étant très grand, le modèle est résolu par une méthode de génération de colonnes. Le sous-problème est un problème de plus court chemin dans un réseau avec des variables de ressource. Les auteurs utilisent un réseau de type lignes de temps augmentées avec les agrégations habituelles proposées par HANE *et al.* (1995), lorsque les vols de repositionnement ne sont pas inclus. S'il le sont, on doit utiliser un réseau de plus grande taille où les connexions sont explicitement représentées pour pouvoir spécifier des coûts propres aux segments de vol et aux connexions. Si les contraintes d'entretien ne requièrent qu'un temps maximal hors de la base d'entretien, sans un nombre d'heures de vol maximal, le réseau peut être modifié pour que le sous-problème ne soit qu'un problème de plus court chemin simple. Les auteurs ont testé leur méthode avec des données provenant d'une compagnie faisant du long courrier et dont les opérations comprennent 1124 segments de vol sur une semaine en visitant 40 villes avec 89 avions de neuf types différents. La résolution séquentielle rend le problème des itinéraires d'avions non réalisable pour cinq des neuf flottes, alors que la résolution intégrée a permis de trouver une solution réalisable dans tous les cas. Un branchement heuristique basé sur une tolérance par rapport à l'écart d'intégrité

rend les temps de calcul raisonnables. La notion de journée de travail utilisée dans l'article peut être très restrictive dans le cas des vols internationaux où les arrêts aux stations d'entretien sont plus rares. Permettre l'entretien pendant la journée ne peut qu'améliorer la solution, mais demande l'utilisation d'une formulation différente et peut être plus difficile à résoudre.

Les auteurs proposent aussi une méthode de génération de colonnes et de coupes pour résoudre le problème des itinéraires d'avions lorsque l'on ajoute la contrainte de l'utilisation égale des avions. Pour ce problème, on suppose que le type d'appareil est fixé et que les vols sont de courte durée dans un réseau de type *hub-and-spoke*. Pour une utilisation égale des avions, on force tous les avions à emprunter le même itinéraire (FEO et BARD (1989), GOPALAN et TALLURI (1998) et TALLURI (1998)). La méthode génère des contraintes d'élimination de sous-tours au besoin. À chaque noeud de branchement, on alterne entre la génération de colonnes et la génération de coupes. On utilise une estimation des coûts réduits pour générer les colonnes car on ne connaît pas la valeur des variables duales des contraintes non-générées.

Certains auteurs proposent d'intégrer l'affectation des types d'appareils et la création des rotations d'équipages. Les contraintes liant les deux problèmes assurent que si un segment de vol est affecté à un type d'avion, alors ce segment doit être affecté à un équipage faisant partie de la même famille d'appareils. BARNHART *et al.* (1998c) proposent une méthode qui intègre le choix des types d'appareils et les rotations d'équipages. Les deux problèmes sont en fait résolus séquentiellement, mais lors de la résolution de l'affectation des types d'appareils, une relaxation du problème des rotations d'équipages est incorporée au modèle. L'idée est de guider le choix des types d'appareils en estimant l'impact que cette affectation aura sur le coût des équipages. La relaxation qui sert à estimer ces coûts est basée sur la méthode proposée précédemment par les auteurs (BARNHART et SHENOI (1998)). La présente méthode résout ensuite le problème des rotations d'équipages par famille d'appareils.

Les problèmes sont résolus par une méthode de génération de colonnes. Les auteurs ont testé leur approche à l'aide de données provenant d'une compagnie faisant du long courrier. Les auteurs n'ont utilisé qu'un seul ensemble de données qui comprenait 964 segments de vol pour 20 avions de deux types différents et 452 vols de repositionnement possibles. Le modèle intégré a permis une réduction de 3,21% des coûts d'équipages par rapport au processus totalement séquentiel. La réduction possible n'était que de 3,51%. Les auteurs soulignent que ce modèle intégré est particulièrement utile dans un contexte international. Le temps permis hors de la base est plus long pour les horaires avec des vols internationaux, donc plus dominant dans les coûts, et la structure de coûts simplifiée de la relaxation utilisée ne comprend justement que ces coûts et les coûts des vols de repositionnement. Par contre, comme le modèle n'intègre pas de contrainte d'entretien, les itinéraires d'avions créés par la suite peuvent ne pas être réalisables, surtout dans un contexte international comme celui-ci.

2.4.2 Intégration complète

Certains auteurs proposent aussi d'intégrer la création des itinéraires d'avions et celle des rotations d'équipages. La principale interaction entre ces deux étapes est un temps de connexion variable. Comme le temps minimal de connexion entre deux vols couverts par le même équipage est réduit si le même avion couvre ces deux vols, l'ensemble des rotations réalisables dépend des itinéraires construits.

KLABJAN *et al.* (2002) proposent une méthode qui trouve les rotations d'équipages d'abord, mais en incluant des contraintes supplémentaires qui assurent la réalisabilité du problème des itinéraires d'avions qui est résolu par la suite. La méthode demeure donc séquentielle, mais elle permet d'améliorer le processus de planification. Ce modèle permet aussi de modifier légèrement l'horaire des vols à l'intérieur de fenêtres de temps dans le but de réduire davantage les coûts liés aux équipages. Les auteurs incluent

la plupart des contraintes de validité des contrats de travail et utilise une fonction de coût complexe. Ils montrent que les crédits de vol diminuent lorsque le temps de connexion diminue (par exemple de 45 à 30 minutes). Réduire le temps de connexion augmente le nombre de rotations possibles, ce qui permet une réduction des coûts. Un temps de connexion réduit n'est possible par ailleurs que si l'équipage demeure à bord du même avion pour couvrir les deux vols. Ce choix impose donc une séquence de vols pour un avion, ce qui risque d'augmenter le nombre d'avions au sol à certains moments. Pour s'assurer que les itinéraires d'avions soient réalisables, les auteurs incluent dans le problème des rotations d'équipages certaines contraintes qui comptent le nombre d'avions disponibles. Pour chaque arc d'attente, une contrainte s'assure qu'à tout moment le nombre d'avions au sol imposé par les connexions courtes ne soit pas supérieur au nombre d'avions provenant de la solution du problème d'affectation des types d'appareils. Toutefois, les itinéraires d'avions peuvent quand même être non réalisables car les contraintes d'entretien ne sont pas incluses. La méthode n'adresse pas ce problème car elle est développée pour les vols de courte durée, passant fréquemment par une station d'entretien. Les fenêtres de temps (± 5 ou 10 minutes) sont modélisées en faisant varier les temps de connexion maximal et minimal. On appelle décalage la différence entre l'heure révisée et l'heure originale. La méthode comprend trois étapes. On génère premièrement, à l'aide d'une procédure rapide, des rotations potentielles qui fixent les décalages de façon à ce que les rotations soient réalisables. On résout ensuite le problème des rotations d'équipages, par un modèle de partitionnement d'ensemble, à partir des rotations générées à l'étape précédente auxquelles on ajoute des contraintes approximatives pour compter les avions. Ces contraintes sont approximatives car les décalages peuvent changer l'ensemble des arcs d'attente. Finalement, si ces contraintes approximatives n'assurent pas des itinéraires respectant la disponibilité des avions, on ajoute des contraintes supplémentaires qui empêchent les groupes de rotations à problèmes d'être choisis ensemble. La méthode a été testée à l'aide de données provenant de United Airlines pour quatre types

d'appareils qui comprenaient entre 100 et 450 segments de vol quotidiens. La méthode a permis de diviser par deux les crédits de vol par rapport à une approche sans considération pour les itinéraires d'avions. Les modifications à l'horaire ont permis une réduction supplémentaire de 25%. Les auteurs rapporte un grand écart d'intégrité, mais ne donne aucune précision quant à la méthode de séparation et d'évaluation utilisée.

CORDEAU *et al.* (2001) présentent un modèle intégrant complètement la création des itinéraires d'avions et des rotations d'équipages. Les auteurs proposent une approche de résolution basée sur la méthode de décomposition de Benders pour des problèmes sur quelques jours comprenant des conditions initiales et finales. Cette méthode de décomposition évite de devoir résoudre un problème qui comprend à la fois des variables correspondant aux rotations d'équipages et des variables correspondant aux itinéraires d'avions. Le processus de résolution itère entre un problème-maître qui résout le problème des itinéraires d'avions, et un sous-problème qui résout celui des rotations d'équipages jusqu'à ce que l'on puisse prouver l'optimalité de la relaxation linéaire. En raison de leur structure particulière, ces problèmes sont eux-mêmes décomposés et résolus par une méthode de génération de colonnes. Les connexions courtes sont fixées par le problème-maître et le sous-problème construit des rotations d'équipages selon les connexions courtes déjà fixées. On se débarrasse ainsi des contraintes liant les deux problèmes. Une méthode de séparation et d'évaluation heuristique est utilisée afin d'obtenir des solutions entières. Les contraintes d'entretien régulier forcent les itinéraires à visiter une des stations d'entretien pendant huit heures, au moins une fois à tous les quatre jours. Il n'y a par contre aucun coût associé aux séances d'entretien. Les contraintes de validité des rotations et des journées de travail comprennent un temps minimal de connexion de 45 minutes (qui peut être réduit à 30 minutes si l'équipage reste sur le même avion), une durée minimale pour les repos de nuit, une limite pour le temps total hors de la base et pour le nombre maximal

d'atterrissages dans une journée de travail. Les séquences combinées sont fournies et les vols de repositionnement sont permis seulement parmi les vols de la compagnie aérienne elle-même. Les auteurs utilisent une fonction de coût non linéaire où, par exemple, les coûts de connexion dépendent du temps d'attente de l'équipage. Les coûts reliés aux vols de repositionnement sont aussi inclus, alors que le minimum garanti pour les équipages ne l'est pas. Les rotations créées peuvent donc être anormalement courtes. Les auteurs ont testé leur méthode avec des données provenant de trois flottes d'une compagnie aérienne canadienne de taille moyenne. La plus grande instance comprend 525 segments de vol sur trois jours avec 35 appareils et 67 équipages. Les auteurs ont résolu les problèmes de routage d'avions et de rotations d'équipages pour toutes les instances d'abord séquentiellement et ensuite de façon intégrée. Il ont utilisé les mêmes méthodes de génération de colonnes pour la résolution séquentielle et pour le problème-maître ou le sous-problème de la décomposition de Benders. Les auteurs ont pu ainsi bien évaluer l'utilité de la résolution simultanée. Les réductions de coûts grâce au modèle intégré sont en moyenne de 9,4% des coûts variables. Comme l'approche simultanée encourage les connexions courtes, qui contribuent à diminuer les coûts, les auteurs ont testé une approche séquentielle qui maximise l'utilisation des connexions courtes lors de la création des itinéraires d'avions. Cette nouvelle approche séquentielle donne des solutions de meilleurs coûts que l'ancienne, mais qui demeurent environ 5,5% plus chères que les solutions fournies par l'approche simultanée, qui semble ainsi mieux choisir les connexions courtes. Leur méthode ne résout, par contre, que des problèmes de taille moyenne.

COHN et BARNHART (2003) présentent aussi un modèle intégré, mais, au lieu d'inclure la formulation du problème de routage d'avions dans le modèle, des variables représentant des solutions complètes au problème d'avions sont utilisées. Ceci réduit considérablement le nombre de contraintes dans le modèle, mais amène un très grand nombre de variables supplémentaires. Dans ce *modèle étendu de*

rotations d'équipages, une solution d'avions qui n'utilise pas au moins une connexion courte n'a pas besoin d'être incluse, car les connexions courtes sont le seul lien entre le problème des itinéraires d'avions et celui des rotations d'équipages. Les auteurs montrent, en fait, que seulement un sous-ensemble des solutions réalisables complètes au problème d'avions doit être inclus dans le modèle, soit une colonne par *ensemble unique et maximal de connexions courtes* (UM). Ces colonnes peuvent être générées individuellement et séquentiellement, lors d'un pré-traitement, en résolvant un ensemble de problèmes de routage d'avions avec des contraintes additionnelles et une fonction-objectif modifiée. Ce problème modifié de routage d'avions est un problème combinatoire difficile, mais les auteurs prétendent que la génération de même un petit sous-ensemble des colonnes UM produit des solutions de très bonne qualité au problème intégré. Ils proposent aussi de résoudre le modèle étendu de rotations d'équipages par une méthode de génération de colonnes, où à la fois les rotations d'équipages réalisables et les solutions complètes d'avions sont générées dynamiquement. Comme la méthode ne permet pas de savoir si la valeur de la solution du modèle étendu de rotations d'équipages est proche de la valeur optimale lorsqu'un sous-ensemble de colonnes UM est utilisé, Cohn et Barnhart proposent aussi une autre approche, appelée *modèle contraint de rotations d'équipages*, afin d'obtenir une mesure de qualité. Avec cette approche, un problème de rotations d'équipages, où toutes les connexions courtes sont permises, est d'abord résolu. Si l'ensemble de connexions courtes utilisé par la solution d'équipages mène à un problème de routage d'avions réalisable, alors une solution optimale au problème intégré a été trouvée. Sinon, une coupe de réalisabilité est introduite dans le problème d'équipages pour interdire un sous-ensemble non réalisable de connexions courtes, et le processus se répète. Les auteurs suggèrent de résoudre parallèlement le modèle étendu de rotations d'équipages et le modèle contraint de rotations d'équipages pour générer, respectivement, des bornes supérieures et des bornes inférieures, et ainsi pouvoir mesurer la qualité des solutions. Les auteurs ont testé le modèle étendu de rotations d'équipages, où les

colonnes UM sont générées dans une étape de pré-traitement, sur de petites instances (125 vols). Ils parlent des gains dû à l'intégration, mais ne donnent aucune information à propos des temps de calculs.

Les caractéristiques des principaux modèles intégrant certaines étapes de la planification sont résumées dans le tableau 2.4.

2.5 L'intégration des tournées de véhicules et des horaires d'équipages en transport en commun urbain

La planification du transport en commun urbain implique la création de tournées pour les véhicules et d'horaires pour les chauffeurs de manière à servir tous les voyages prévus à coût minimal. La procédure habituelle résout ces deux problèmes séquentiellement, c'est-à-dire que les autobus sont premièrement affectés aux voyages, et ensuite les chauffeurs sont affectés aux autobus. Comme les coûts des chauffeurs dominant habituellement les coûts des véhicules en transport urbain, l'intégration de ces deux étapes de la planification est avantageuse globalement. Le problème intégré est, par contre, de plus grande taille et inclut des contraintes qui lient les véhicules aux chauffeurs. Par exemple, comme les chauffeurs ne peuvent changer de véhicule ailleurs qu'à quelques endroits spécifiques, certaines séquences de voyages affectées à un chauffeur doivent être affectées à un même véhicule. Même le problème intégré qui ne comprend qu'un dépôt et un type de véhicule ne peut être résolu en un temps polynomial étant données les contraintes complexes provenant des contrats de travail liant les chauffeurs aux compagnies de transport. Nous allons présenter les principales méthodes retrouvées dans la littérature pour résoudre ce problème intégré simplifié.

Plusieurs heuristiques ont été proposées pour résoudre ce problème, comme par exemple celles de BALL *et al.* (1983), TOSINI et VERCELLIS (1992) et PATRIKALAKIS

TAB. 2.4 Caractéristiques des principaux modèles intégrant des étapes de la planification

Auteurs	Type de problème	Modèle	Fonction objectif	Contraintes supplém.	Approche de résolution	Taille des probl.
BARNHART et al. (1998)	Appareils/ itinéraires Long courr. Entretien de nuit	Journées de travail augmentées	Coûts d'opération des avions Revenus perdus Séq. combin.	Entretien rég : 4 jours	Gén. de col. Branchement heuristique	1124 vols 9 types 89 avions
BARNHART et al. (1998)	Appareils/ rotations	Appareils avec relax. équipages	Coûts d'opération des avions Revenus perdus Temps total hors base	Lien appareil et équipage Temps de connexion min Repos de nuit	Gén. de col.	964 vols 2 types
KLABJAN et al. (2002)	Itinéraires/ rotations Court courr. Quotidien Fenêtres de temps	Équipages avec dispon. des avions	Crédits de vol	Temps min/max de connexion Temps min/max de repos de nuit 8 hres en 24 Temps hors base	Génération de rotations admissibles PL	450 vols
CORDEAU et al. (2001)	Itinéraires/ rotations Moyen courr. Conditions initiales et finales	Intégré	Linéaire par morceaux Connexions Repos de nuit Deadheads	Entretien rég : 4 jours Temps min/max de connexion Repos de nuit Temps hors base Lien avions et équipages avec connexions courtes	Décomposition de Benders Branchement heuristique	525 vols 67 équip.
COHN ET BARNHART (2001)	Itinéraires/ rotations	Intégré Équipages et solutions complètes d'avion	Coûts d'équipages	Lien avions et équipages avec connexions courtes	Gén. de col. Branchement heuristique Génération de coupes	125 vols

et XEROCOSTAS (1992). Certaines formulations exactes résolues par une approche possiblement heuristique ont aussi été suggérées. FRELING *et al.* (1995) décrivent une formulation contenant trois composantes : a) une structure d'affectation pour les tournées de véhicules, b) des contraintes de partitionnement d'ensemble pour les horaires des équipages et c) une série de contraintes liant le problème des véhicules au problème des équipages. Les auteurs présentent une méthode de relaxation lagrangienne qui dualise les contraintes liantes. Le sous-problème se décompose en un problème pour les véhicules et un second pour les équipages. Les horaires possibles pour les chauffeurs (ceux qui respectent les règles du contrat de travail) peuvent être fournis a priori ou créés par une méthode de génération de colonnes. La suite de ce travail de recherche est présentée par FRELING (1997). L'auteur y décrit une procédure heuristique pour trouver efficacement des solutions entières au problème. Les auteurs ont testé leur méthode à l'aide de problèmes générés aléatoirement comprenant jusqu'à 120 voyages et aussi à l'aide de problèmes réels comprenant jusqu'à 148 voyages. L'objectif était de minimiser premièrement le nombre d'autobus et de chauffeurs requis et ensuite les coûts d'opération. La méthode a trouvé des solutions réalisables dont l'écart d'optimalité se situe entre 6% et 26% pour les problèmes aléatoires et entre 0% et 7,6% pour les problèmes réels.

La première approche de résolution exacte a été proposée par HAASE et FRIBERG (1999). Les auteurs utilisent une formulation de partitionnement d'ensemble où une variable représente un horaire pour un équipage ou une tournée pour un véhicule. Des contraintes supplémentaires sont ajoutées pour lier les deux types de variables. Un algorithme de séparation, d'évaluation et de coupes (*branch-and-price-and-cut*) est utilisé pour résoudre le problème. Les tournées et les horaires sont créés par génération de colonnes et les coupes de cliques proposées par HOFFMAN et PADBERG (1993) sont considérées pour les véhicules. Seuls des petits problèmes qui comprennent jusqu'à 20 voyages ont pu être résolus à l'optimalité par la méthode.

HAASE *et al.* (2001) présentent une approche de résolution pour le problème des équipages incluant des considérations pour le problème des autobus. Les auteurs utilisent une formulation de partitionnement d'ensemble avec des contraintes supplémentaires pour les véhicules. Une méthode de séparation et d'évaluation, où les relaxations linéaires sont résolues par une méthode de génération de colonnes, est utilisée. Les auteurs montrent que les contraintes supplémentaires assurent de pouvoir trouver par la suite, en un temps polynômial, une solution optimale pour le problème des véhicules. Leur méthode a été testée avec des données générées aléatoirement qui comprenaient entre 40 et 140 voyages. Les petits problèmes ont presque tous été résolus à l'optimalité, alors qu'environ la moitié des problèmes de plus grande taille ont pu l'être.

Plus récemment, HUISMAN *et al.* (2005) présentent deux modèles et algorithmes différents pour l'intégration des tournées de véhicules et des horaires d'équipages, dans le cas où il y a plusieurs dépôts. Les deux approches combinent une méthode de génération de colonnes à une méthode de relaxation lagrangienne. En fait, le problème-maître de génération de colonnes est résolu par relaxation lagrangienne. La différence entre les deux méthodes réside principalement dans la manière de calculer la borne inférieure. La deuxième méthode se base sur un modèle contenant uniquement des variables d'équipages. Les routes d'autobus peuvent être obtenues implicitement à partir des horaires d'équipages. Les auteurs ont testé leurs méthodes sur des données réelles provenant de la plus grande entreprise de transport par autobus des Pays-Bas et aussi sur des données qu'ils ont générées aléatoirement. Les résultats numériques montrent que les instances de taille moyenne peuvent être résolues par les méthodes proposées et que pour les plus grandes d'entre elles, la méthode se basant sur un modèle contenant des variables pour les équipages ainsi que des variables pour les véhicules est meilleure.

CHAPITRE 3

ORGANISATION DE LA THÈSE

Comme le montre la revue de la littérature, les différents problèmes de la planification tactique en transport aérien ont traditionnellement été résolus de manière séquentielle, et ce, pour en faciliter la résolution. Étant donné le potentiel de gains pour les compagnies aériennes, les travaux récents s'orientent plutôt vers l'intégration, soit partielle ou complète, des différentes étapes de la planification. L'objectif principal de cette thèse est de développer des modèles et des approches de résolution pour traiter des problèmes intégrés d'itinéraires d'avions et de rotations d'équipages de grande taille.

Quand une connexion entre deux segments de vol est tout juste trop longue pour être considérée comme courte, un équipage peut être affecté aux deux vols de la connexion même s'ils ne sont pas effectués par le même appareil. Cependant, comme ce type de connexion peut provoquer des délais (lorsqu'il y a du retard au premier segment de vol), les compagnies aériennes veulent habituellement en réduire le nombre. Tenir compte de cette contrainte complique le problème intégré d'avions et d'équipages mais cela rend les solutions produites plus robustes. Une solution est considérée plus robuste quand une perturbation à l'horaire est moins susceptible de retarder les vols suivant la perturbation.

Nous proposons au chapitre 4 une approche qui introduit une formulation générale pour le problème intégré d'avions et d'équipages donnant des solutions plus robustes. Résoudre le modèle général intégré sans décomposition est difficile, même lorsque

le nombre de segments de vol est petit. En utilisant la décomposition de Benders, la formulation du problème intégré peut être décomposée de façon à obtenir deux problèmes qui sont plus faciles à résoudre. La méthodologie présentée dans ce chapitre montre qu'un choix judicieux des variables demeurant dans le problème-maître de Benders peut faciliter la résolution du problème. La plupart des coûts du problème intégré sont attribuables aux rotations d'équipages et mettre le problème d'avions dans le sous-problème de Benders fait en sorte que celui-ci transfère principalement de l'information sur sa réalisabilité au problème-maître et très peu d'information d'optimalité (de coûts). Comme il peut exister plus d'une solution au sous-problème dual à chaque itération de la méthode de décomposition de Benders, même si chacune de ces solutions permet la construction d'une coupe d'optimalité valide, certaines peuvent mener à des coupes plus fortes que d'autres. Une autre contribution de ce chapitre est de montrer, dans le cas où le problème-maître est résolu par génération de colonnes, comment identifier une solution optimale au sous-problème dual qui mène à une coupe Pareto-optimale (une coupe qui n'est dominée par aucune autre).

Lorsque l'on considère un niveau supplémentaire d'intégration en permettant de prendre certaines décisions par rapport à l'horaire des vols, la qualité globale des solutions peut être améliorée. Un modèle intégré d'avions et d'équipages tire davantage profit de cette nouvelle flexibilité car les heures de départ sont modifiées en tenant compte de l'effet, non seulement sur les itinéraires d'avions, mais aussi sur les horaires d'équipages et cela serait impossible avec une résolution séquentielle où les modifications faites à un niveau de décisions donné pourrait avoir des conséquences fâcheuses au niveau suivant. Évidemment, il faut ajouter des contraintes afin de s'assurer que les mêmes heures de départ soient choisies à la fois pour les itinéraires d'avions et pour les rotations d'équipages, ce qui ajoute un degré de difficulté à la résolution du problème.

Une extension directe de la méthode présentée au chapitre 4 ne peut toutefois résoudre le problème intégré avec modifications à l'horaire que sur des instances de petite taille. Certains raffinements doivent y être ajoutés et nous les présentons au chapitre 5. Le modèle intégré avec modifications à l'horaire contient un grand nombre de contraintes liantes de connexion courte qui rendent le problème difficile à résoudre (chaque vol formant une connexion a plusieurs horaires possibles). On peut diminuer le nombre de ces contraintes en les agrégeant de façon à ne conserver qu'une contrainte liante par connexion courte. En fait, nous proposons dans ce chapitre une reformulation plus compacte du problème et nous montrons que non seulement la formulation agrégée est équivalente à la formulation complète en nombres entiers, mais que certaines relaxations linéaires mixtes le sont elles aussi. Nous proposons aussi une méthode de génération dynamique de coupes (à la fois dans le problème-maître de Benders et dans le sous-problème de Benders) qui permet de résoudre efficacement des problèmes de grande taille. La dernière contribution de ce chapitre est de montrer, empiriquement, que le modèle avec modifications à l'horaire produit des solutions entières à la fois moins chères pour ce qui est des coûts d'équipages et nécessitant un plus petit nombre d'appareils par rapport à un modèle avec un horaire fixe.

Le chapitre 6 étudie en détail le modèle intégré avions-équipages de base, soit celui sans possibilité de modifications à l'horaire et sans pénalité attribuée aux solutions moins robustes. Lorsque la décomposition de Benders est utilisée, avec en problème-maître la création des rotations d'équipages, le sous-problème de Benders (avions) ne génère, en fait, que des coupes de réalisabilité à ajouter au problème-maître à chaque itération de l'algorithme. Il a aussi été proposé, dans la littérature, de résoudre le problème avions-équipages de base, itérativement, à l'aide d'un problème de rotations d'équipages modifié, en ajoutant des coupes de réalisabilité provenant d'une famille de coupes prédéfinie, jusqu'à ce que l'ensemble des connexions courtes utilisé par les équipages dans la solution courante soit réalisable pour les avions. Dans ce chapitre,

nous présentons une comparaison théorique des coupes de Benders et des autres familles de coupes de réalisabilité pour le problème avions-équipages de base. Nous montrons que toutes les coupes provenant d'une famille prédéfinie sont incluses dans l'ensemble des coupes de Benders, et qu'elles peuvent aussi être dominées par ces dernières.

La discussion inclut aussi une comparaison des coupes de réalisabilité et des coupes d'optimalité de Benders à travers la description du polyèdre de leur sous-problème dual respectif. Les coupes d'optimalité de Benders sont générées à partir du problème de routage d'avions rendu réalisable par l'introduction de variables artificielles. Le choix des variables artificielles a un impact sur le polyèdre du sous-problème dual, donc sur les coupes générées. Nous montrons qu'en ne mettant qu'une seule variable artificielle dans le sous-problème d'avions (donc une seule contrainte dans le sous-problème dual), il y a une correspondance exacte entre les points extrêmes du sous-problème dual borné et les rayons extrêmes du sous-problème dual d'origine. Par ailleurs, nous montrons que certaines combinaisons de rayons extrêmes donnent des coupes de réalisabilité plus fortes que celles générées directement à partir des rayons extrêmes. Une autre contribution de ce chapitre est de proposer une méthode de renforcement simple qui peut être appliquée à tous les types de coupes de réalisabilité pour le problème intégré avions-équipages de base.

Finalement, la dernière partie de la thèse comporte une discussion générale et une conclusion.

CHAPITRE 4

A COMPUTATIONAL STUDY OF BENDERS DECOMPOSITION FOR THE INTEGRATED AIRCRAFT ROUTING AND CREW SCHEDULING PROBLEM

Article écrit par Anne Mercier, Jean-François Cordeau et François Soumis, publié dans *Computers & Operations Research* 32, pages 1451-1476, 2005.

Le problème intégré de la création d'itinéraires d'avions et d'horaires d'équipages consiste à déterminer un ensemble de routes d'avions et de rotations d'équipages tels que chaque vol est couvert par un appareil et un équipage, tout en satisfaisant un ensemble de contraintes supplémentaires. Alors que certaines contraintes supplémentaires ne concernent que les équipages ou les avions, des contraintes liantes imposent des temps minimums de connexion pour les équipages qui dépendent des connexions utilisées par les appareils. La connexion entre deux segments de vol est dite *courte* si elle est réalisable mais que sa durée est plus petite que le temps de connexion minimal pour les équipages. Dans ce cas, les deux segments de vol ne peuvent être couverts en séquence par le même équipage que si les vols sont aussi couverts par le même appareil. Sinon, la durée de la connexion est insuffisante pour que l'équipage puisse faire le changement d'appareil.

Un modèle de couverture de tâches (vols) avec des contraintes liantes de connexions courtes est utilisé pour modéliser le problème intégré. Une contrainte supplémentaire limite le nombre d'avions disponibles. Chaque itinéraire d'avion doit, par ailleurs, respecter une limite sur le nombre d'heures en vol et sur le nombre de jours séparant deux visites à une station d'entretien. Chaque journée de travail dans une rotation

d'équipage doit respecter une limite sur sa durée totale, sur le nombre d'heures en vol ainsi que sur le nombre d'atterrissages. De plus, le nombre de journées de travail dans une rotation ne doit pas dépasser une certaine limite. Ces contraintes de réalisabilité des chemins (itinéraires d'avions ou rotations d'équipages) sont modélisées avec l'utilisation de ressources et on peut les traiter directement avec un algorithme de programmation dynamique dans le cadre de la génération de colonnes. Un réseau espace-temps est utilisé pour générer les chemins d'avions et les chemins d'équipages.

Nous présentons ici une approche qui introduit une formulation générale pour le problème intégré d'avions et d'équipages. Quand une connexion entre deux segments de vol est tout juste trop longue pour être considérée comme courte, un équipage peut être affecté aux deux vols de la connexion même s'ils ne sont pas couverts par le même appareil. Cependant, comme ce type de connexion peut provoquer des retards, les compagnies aériennes veulent habituellement en réduire le nombre. Tenir compte de cette contrainte complique le problème intégré d'avions et d'équipages mais cela rend les solutions produites plus robustes. Une solution est considérée plus robuste quand une perturbation à l'horaire est moins susceptible de retarder les vols suivant la perturbation.

La connexion entre deux segments de vol est dite *restreinte* si elle est réalisable et que sa durée est plus grande que le temps de connexion minimal pour les équipages, mais plus petite qu'un seuil donné qui correspond à un temps de connexion *idéal*. Si une connexion est restreinte, les deux segments de vol peuvent être couverts en séquence par le même équipage mais, si les vols ne sont pas aussi couverts par le même appareil, des délais risquent d'être introduits à l'horaire des vols. En incluant un terme de pénalité pour ce type de connexion dans la fonction-objectif, la formulation générale produit des solutions plus robustes.

Résoudre le modèle général intégré sans décomposition est difficile, même lorsque le nombre de segments de vol est petit. En utilisant la décomposition de Benders,

la formulation du problème intégré peut être décomposée de façon à obtenir deux problèmes qui sont plus faciles à résoudre. En fait, cette formulation intégrée inclut à la fois des variables d'équipages et d'avions, et la décomposition de Benders peut être utilisée pour reformuler le problème afin de réduire le nombre de variables en augmentant cependant le nombre de contraintes. Les contraintes supplémentaires peuvent toutefois être générées dynamiquement lorsqu'elles sont violées.

La plupart des coûts du problème intégré sont attribuables aux rotations d'équipages. En effet, si l'on fait abstraction des pénalités rattachées aux connexions restreintes, le problème de routage d'avions devient un problème de réalisabilité. Inverser l'ordre naturel de résolution (l'ordre habituellement choisi par les approches séquentielles) dans la méthode de décomposition de Benders, en mettant le problème d'avions dans le sous-problème de Benders, fait en sorte que celui-ci transfère principalement de l'information sur sa réalisabilité au problème-maître et très peu d'information d'optimalité (de coûts). En choisissant le bon ordre de résolution, le nombre de coupes de Benders qui doivent être générées diminue, en effet, significativement.

Un algorithme en trois phases est utilisé pour résoudre le problème intégré. Dans la phase I, toutes les contraintes d'intégrité sont relaxées et la relaxation linéaire est résolue à optimalité par une méthode de décomposition de Benders dans laquelle à la fois le problème-maître d'équipages et le sous-problème d'avions sont résolus par une méthode de génération de colonnes. En conservant toutes les coupes de Benders générées à la phase I, la phase II réintroduit les contraintes d'intégrité pour les variables du problème-maître (chemins d'équipages) et résout le problème résultant en générant des coupes additionnelles. Dans cette phase, le problème-maître en nombres entiers doit être résolu à chaque itération de l'algorithme de décomposition de Benders. Comme les contraintes liantes de connexions courtes dans le sous-problème primal forcent un avion à faire celles qui sont utilisées par les équipages dans le problème-maître, il est possible que le sous-problème primal en nombres entiers soit non-

réalisable à la phase III (pour les chemins d'équipages fixés par le problème-maître) même si le problème d'origine est réalisable. Une étape est donc ajoutée après la troisième phase pour vérifier la réalisabilité du sous-problème en nombres entiers et, si nécessaire, revenir à la phase II pour résoudre de nouveau le problème-maître en nombres entiers avec une contrainte supplémentaire qui empêche le même sous-ensemble de connexions courtes d'être choisi par les équipages.

Comme il est possible que le sous-problème primal de Benders soit dégénéré, il peut exister plus d'une solution au sous-problème dual à chaque itération de la méthode de décomposition de Benders. Même si chacune de ces solutions (points extrêmes) permet la construction d'une coupe d'optimalité valide, certaines peuvent mener à des coupes plus fortes que d'autres. Pour identifier une solution optimale au sous-problème dual qui mène à une coupe Pareto-optimale (une coupe qui n'est dominée par aucune autre), on doit résoudre un sous-problème dual auxiliaire après chaque résolution du sous-problème de Benders. La fonction-objectif du problème auxiliaire compare toutes les coupes possibles à un point de l'intérieur relatif de la région admissible du problème-maître. De cette façon, on essaie de choisir parmi les points extrêmes optimaux celui qui donne une contrainte qui coupe le plus la région admissible du problème-maître. Comme le problème-maître de Benders est résolu par génération de colonnes dans la méthode en trois phases, son ensemble de variables n'est pas entièrement connu. Une représentation explicite du polyèdre n'est donc pas disponible et trouver un point de l'intérieur relatif de la région admissible du problème-maître devient une tâche plus difficile. Bien que cela puisse altérer la force de la coupe, choisir un point en dehors de la région admissible n'empêche pas le problème auxiliaire de générer une coupe valide pour le problème-maître. Fixer arbitrairement les coefficients des variables de la fonction-objectif du sous-problème auxiliaire est une alternative efficace qui permet de générer des coupes de Benders valides et fortes.

La méthode en trois phases appliquée à la formulation générale du modèle intégré, et ses raffinements, ont été testés avec des données provenant de huit flottes de deux compagnies aériennes de grande taille (jusqu'à 700 vols par jour pour une flotte). Lorsqu'elle est comparée, avec les mêmes données, à la méthode où le problème-maître de Benders résout le problème de routage d'avions, l'approche proposée diminue par un facteur de 7, en moyenne, le temps nécessaire pour résoudre la formulation générale. La différence quant à la performance est principalement attribuable à un plus petit nombre de coupes de Benders générées. Lorsque des coupes Pareto-optimales sont générées, le temps CPU total est diminué davantage. Globalement, les deux raffinements accélèrent en moyenne la vitesse de résolution de la méthode de décomposition de Benders par un facteur de plus de 20 et ce, sans détériorer la qualité des solutions.

A Computational Study of Benders Decomposition
for the Integrated Aircraft Routing and Crew
Scheduling Problem

ANNE MERCIER

École Polytechnique de Montréal

JEAN-FRANÇOIS CORDEAU

HEC Montréal

FRANÇOIS SOUMIS

École Polytechnique de Montréal

December 2003

Abstract

The integrated aircraft routing and crew scheduling problem consists in determining a minimum-cost set of aircraft routes and crew pairings such that each flight leg is covered by one aircraft and one crew, and side constraints are satisfied. While some side constraints involve only crews or aircraft, linking constraints impose minimum connection times for crews that depend on aircraft connections. We propose an enhanced model incorporating robustness to handle these linking constraints and compare two Benders decomposition methods - one with the aircraft routing problem as the master problem and one with the crew pairing problem. We also study the impact of generating Pareto-optimal cuts on the speed of convergence of these methods. Computational experiments performed on test instances provided by two major airlines show that the proposed approach yields high quality solutions in reasonable computing times.

Keywords: aircraft routing; crew scheduling; integrated planning; Benders decomposition; column generation.

Introduction

After creating a schedule that defines origin and destination cities as well as departure and arrival times for each flight leg to be flown during a given period, most airlines use a sequential procedure to plan their operations. The first step of this procedure is the *fleet assignment problem* which consists in assigning an aircraft type to each flight leg so as to maximize anticipated profits. For each aircraft type, an *aircraft routing problem* is then solved to determine the sequence of flight legs to be flown by each individual aircraft so as to cover each leg exactly once while ensuring appropriate aircraft maintenance. To comply with safety regulations stipulated by transport authorities, airlines must ensure that every aircraft regularly undergoes different types of maintenance checks that vary in scope, frequency and duration. For example, routine checks are usually performed every 3 to 4 days whereas more complete checks performed annually may require the aircraft to be grounded for several days at once. When solving the fleet assignment problem and the aircraft routing problem, airlines sometimes take into consideration *through values* that represent the extra revenues obtained by assigning the same aircraft to a pair of consecutive flight legs (i.e., a *through*) so that passengers flying from the origin of the first leg to the destination of the second leg do not have to change aircraft. When through values are not considered, the aircraft routing problem usually reduces to a feasibility problem.

Given fixed aircraft routes and a set of work rules defined by the collective agreement, the airline then builds crew rotations or *pairings* by solving a *crew scheduling problem*. In general terms, a crew pairing is a sequence of duty and rest periods that typically lasts between two and five days. A *duty period* corresponds to a single workday for a crew and can be seen as a sequence of flight legs separated by short rest periods, called *sits*. A pairing is thus a sequence of duty periods separated by overnight rests. Every pairing begins and ends at a specific location, called a *crewbases*, and must satisfy a

set of applicable work rules related to a large number of factors such as flight time, rest time, connection time, etc. The objective of the crew scheduling problem is to determine a minimum-cost set of pairings so that every flight leg is assigned a qualified crew. The cost of a pairing depends not only on the total flight time but also on the waiting time during connections as well as on related accommodation expenses when overnight connections take place outside the crewbase. In the last step of the planning process, pairings are finally combined to form personalized monthly schedules that are assigned to employees by solving a *crew bidding problem* or a *crew rostering problem*.

Using a sequential procedure considerably reduces the complexity of the process but may also yield solutions that are far from optimal, as demonstrated by CORDEAU *et al.* (2001) and by COHN and BARNHART (2003). The main source of interaction between aircraft routing and crew scheduling resides in variable connection times. The minimum connection time (or *sit time*) required between two successive flight legs covered by the same crew depends on whether the same aircraft is used on both legs. The set of feasible pairings therefore depends on the aircraft routing decisions made in the previous step. Hence, a suboptimal solution is likely to be obtained if a sequential planning procedure is used.

A connection that is too short to be made by a crew when the two associated legs are not flown by the same aircraft is said to be *short*. When a connection between two flight legs is just long enough not to be considered short, a crew can thus be assigned to the two legs even if they are not flown by the same aircraft. However, since these connections are likely to cause delays, airlines usually wish to restrict their use. Addressing this issue complicates the integrated aircraft routing and crew scheduling problem but it also makes its solution more robust. A solution is said to be more robust when disruptions are less likely to delay the flights following the unforeseen event.

Several modeling and solution approaches have been proposed to individually address the aircraft routing and crew pairing problems. The former problem was addressed, among others, by DASKIN and PANAYOTOPOULOS (1989), FEO and BARD (1989), CLARKE *et al.* (1997), GOPALAN and TALLURI (1998) and TALLURI (1998). Numerous contributions regarding the different variants of the crew scheduling problem can also be found in the operations research literature. For an overview, the reader is referred to the recent survey of BARNHART *et al.* (2003).

Until recently, the integration of aircraft routing and crew scheduling decisions had, however, received little attention in the operations research literature. CORDEAU *et al.* (2001) presented a model that integrates the complete aircraft routing and crew pairing formulations to which is added one linking constraint for each short connection. To handle these linking constraints, a solution approach based on Benders decomposition is proposed. The solution process iterates between a master problem that solves the aircraft routing problem, and a subproblem that solves the crew pairing problem. Short connections are fixed by the master problem and the subproblem constructs minimum cost crew pairings using only the fixed set of short connections. Because of their particular structure, both of these problems are solved by column generation. A heuristic branch-and-bound method is used to obtain integer solutions. On a set of test instances based on data provided by a Canadian airline, the integrated approach produced average cost savings of 9.4% in comparison with the sequential planning process commonly used in practice. The largest instance solved contains more than 500 flight legs over a three-day period. The authors therefore demonstrated the usefulness of solving an integrated model, but their method can only solve medium-size instances in a reasonable amount of computation time.

KLABJAN *et al.* (2002) described a partially integrated approach that solves the crew scheduling and aircraft routing problems sequentially but adds plane count constraints to the crew scheduling model. Under the assumption that maintenance is performed

at night when all aircraft are on the ground, these constraints ensure the feasibility of the resulting aircraft routing problem. In addition, the model allows the departure time of each flight leg to be moved within a given time window so as to further reduce crew costs. On test instances involving up to 450 flight legs, this approach produced very significant savings. It is, however, less likely to yield a feasible aircraft routing problem in an international context where maintenance does not necessarily take place at night.

Very recently, COHN and BARNHART (2003) also presented an integrated model, but instead of incorporating the aircraft routing formulation in the model, variables representing complete solutions to the aircraft routing problem are used. This obviously reduces the number of constraints, but may lead to a very large number of additional variables. In this *extended crew pairing* model, an aircraft solution that does not use at least one short connection does not need to be included, since short connections are the only link between the two problems. The authors show that only a subset of the feasible aircraft routing solutions needs to be included in the model, i.e., one column for each *unique and maximal maintenance-feasible short connection set (UM)*. These columns can be generated individually and sequentially, in a preprocessing step, by solving a series of aircraft routing problems with additional constraints and a modified objective function. This modified aircraft routing problem is a difficult combinatorial problem, but the authors claim that generating even a small subset of UM columns can yield high quality solutions. They also propose to solve the extended crew pairing problem by a branch-and-price algorithm in which both crew pairings and aircraft routing solutions are generated dynamically. Although the method works very well for some families of instances, our computational experience has shown that there are cases where generating even a small number of UM columns is difficult. In addition, if only a subset of UM columns is included, there is no guarantee that the value of the extended crew pairing problem will be close to optimality. To

obtain a measure of solution quality, Cohn and Barnhart have also proposed another approach, called the *constrained crew pairing* model. With this approach, a crew pairing problem in which all short connections are permitted is first solved. If the set of short connections used in the solution to this crew pairing problem leads to a feasible aircraft routing problem, an optimal solution has been identified. Otherwise, a cut is introduced in the crew pairing problem to forbid a minimally infeasible subset of short connections and the process is repeated. The extended crew pairing and constrained crew pairing models can be solved in parallel to generate upper and lower bounds, respectively, and thus control solution quality.

Issues related to the introduction of maintenance and crew considerations in the fleet assignment problem were discussed by CLARKE *et al.* (1996), SUBRAMANIAN *et al.* (1994) and RUSHMEIER and KONTOGIORGIS (1997). Another effort in the direction of integration is the work of BARNHART *et al.* (1998c) who proposed an integrated but approximate model for combined fleet assignment and crew scheduling. The approach does not truly solve the two problems simultaneously but rather incorporates a relaxation of the crew scheduling problem in the fleet assignment model. Finally, other interesting contributions with respect to the integration of the planning process are the approaches presented by BARNHART *et al.* (1998a) and DESAULNIERS *et al.* (1997b) for the combined fleet assignment and aircraft routing problem.

Some work has also been done to address the issue of robustness. For the crew scheduling problem, EHRGOTT and RYAN (2002) proposed a bi-criteria approach minimizing crew costs while maximizing robustness. By separating the two conflicting objectives, the authors avoid having to put a price on robustness. Instead, they measure it with scheduled ground time. Their results show significant gains in robustness with only a small increase in costs. CHEBALOV and KLABJAN (2006) present a model that produces more robust crew schedules by including in the

objective function a term maximizing the number of move-up crews, i.e., crews that can potentially be swapped in operations.

In this paper, we assume that the fleet assignment problem has been solved so that the type of aircraft assigned to each flight leg is known. For ease of exposition, we also assume that crews are qualified to fly a single type of aircraft, although this assumption is easily relaxed. In this context, the integrated aircraft routing and crew scheduling problem decomposes into one problem for each aircraft type. Given a set of flight legs to be flown by the aircraft of a specific type, the problem is then to determine a minimum-cost set of aircraft routes and crew pairings such that each flight leg is covered by one aircraft and one crew, and side constraints are satisfied. Our formulation addresses the daily problem, which is common in the crew scheduling literature.

The contribution of this paper is threefold. First, we present a formulation that generalizes upon previous work by introducing the concept of *restricted connections*. Restricted connections designate connections between two flight legs not flown by the same aircraft and just long enough not to be considered short. By including a penalty term for this type of connections in the objective function, the proposed formulation is likely to produce more robust solutions. The second contribution is to compare two implementations of the Benders decomposition approach for the generalized formulation and to introduce computational refinements to improve their speed of convergence. This comparison shows that reversing the order of the solution sequence yields important improvements over the approach of CORDEAU *et al.* (2001). Lastly, we show that on the integrated model excluding restricted connection penalties, the improved Benders decomposition approach outperforms a straightforward implementation of the extended crew pairing approach recently proposed by COHN and BARNHART (2003).

The remainder of the article is organized as follows. The next section introduces some notation and a mathematical formulation of the problem while Section 4.2 presents the solution methodology. Computational experiments that show the efficiency of the proposed refinements are reported in Section 4.3. Conclusions and directions for future work are discussed in the final section.

4.1 Mathematical Formulation

In the daily problem, aircraft routes and crew pairings (that can last more than one day) must be constructed so that each flight leg is covered by one aircraft and one crew, and the solution repeats daily. The number and the duration of the routes and pairings determine the required number of aircraft and crews. For instance, an aircraft route that spans a period of three days will require three different aircraft because each leg in the route must be covered daily. Although most airlines operate according to a schedule that varies throughout the week, the daily problem is usually solved to get an initial solution that will then be refined by taking exceptions into account.

The integrated problem is modeled with a path formulation, as in CORDEAU *et al.* (2001). Consider a set L of flight legs to be flown by a single aircraft type. Each flight leg $l \in L$ is defined by origin and destination stations, and fixed departure and arrival times. Also consider the sets $D = \{1, \dots, d\}$ and $Y = \{1, \dots, y\}$, where d is the maximum number of days separating two visits at a maintenance station and y is the maximum number of days in a crew pairing. Since the daily problem is cyclic, all legs that start from a certain station can be reached from any other leg that ends at that station. This cyclical aspect increases the difficulty of the problem. Nevertheless, the d -day maintenance requirements make it possible to formulate the daily aircraft

routing problem on an acyclic network by including d copies of each leg in the model (or $d - 1$ copies for a leg spanning two days). The solution is still on a single day since all legs are covered exactly once, but the paths generated to construct this solution come from d -day networks. It is reasonable to do so since the maximum number of days between two maintenance checks is usually small. One can also formulate the crew scheduling problem in the same fashion when dealing with domestic flights since the maximum number of days in a pairing is then typically not greater than four.

When d and y differ, distinct network representations must be used to generate the aircraft and the crew paths.

4.1.1 Aircraft routing networks

Let $G^F = (N^F, A^F)$ be a time-space aircraft network where N^F is the node set and A^F is the arc set. Each node $i \in N^F$ corresponds either to the departure of a flight leg $l_i \in L$ on a day $d_i \in D$, or to its arrival. There are two types of arcs: flight arcs and connection arcs. A flight arc is defined between each pair of nodes representing the departure and the arrival of a flight leg on a given day. A connection arc is defined between the arrival node of leg l_i on day d_i and the departure node of leg l_j on day d_j if the destination station of l_i and the origin station of l_j are the same and if the connection time is larger than a given station-specific threshold that represents the minimum connection time for both legs to be covered by the same aircraft. This minimum connection time is often called the *minimum plane turn time* and accounts, among other things, for refueling. Because each leg repeats daily, connections longer than 24 hours (plus the minimum turn time) can be removed without restricting the problem. Hence, every leg arriving at a certain station is connected exactly once to every other leg starting from the same station.

Let h be the time required to perform a regular maintenance check. Let also M and T represent the set of possible starting locations and starting times for paths, respectively. The cardinality of M is the number of maintenance stations, m , which is usually small with respect to the total number of stations. Maintenance centers are normally attached to stations of the physical network and are thus directly accessible at the end of some legs. The cyclical aspect of the problem allows the paths to start and end at a maintenance station without loss of generality. We may thus assume that the maintenance is performed at the end of each path. But since a path can start and end at a different maintenance station and since a complete aircraft route can last more than d days, paths in the acyclic network G^F will represent aircraft routes between maintenance checks (of length not greater than d days) that will be combined to form a valid solution to the daily aircraft routing problem. To do so, additional constraints correctly matching the paths are required. These constraints will be described in Section 4.1.3.

To reduce the cardinality of T (and thus the number of matching constraints), one may discretize the day by dividing it into disjoint but consecutive subperiods of equal length, p . To simplify the notation, we assume here that h is an integer multiple of p . For each subperiod t , let o_t and e_t denote the start and the end of this subperiod, respectively. Using a discretized planning horizon introduces a small error in enforcing the maintenance requirements since an aircraft arriving at a maintenance station in subperiod t cannot be used again before the beginning of subperiod $t + 1 + h/p$, even if it arrived at the maintenance station as early as at time o_t . In this case, the aircraft has to wait more than the required maintenance time, h , before flying again. This error (at most p) decreases as the cardinality of T increases. For example, using four subperiods gives a maximum error of six hours while using twelve decreases this maximum error to two hours. All the experiments reported in this paper use

twelve time periods. For most practical applications, this approximation error can be tolerated.

For each maintenance station $m \in M$ and for each starting time $t \in T$, let s_{mt}^F and q_{mt}^F be source and sink nodes that represent, respectively, the beginning and the end of a route. The source node s_{mt}^F is linked to nodes representing the departure of legs that can be covered directly from station m and that start between o_t and e_t . Nodes that represent the arrival of legs that end at station m before or at o_t and after e_t are linked to the destination node q_{mt}^F . One then defines a network $G_{mt}^F = (N_{mt}^F, A_{mt}^F)$ where $N_{mt}^F = N^F \cup \{s_{mt}^F\} \cup \{q_{mt}^F | m \in M, t \in T\}$. In addition to the arcs from A^F , set A_{mt}^F contains all arcs originating from the source node s_{mt}^F as well as all arcs finishing at a sink node.

4.1.2 Crew pairing networks

Let $G^K = (N^K, A^K)$ be a time-space crew network where N^K is the node set and A^K is the arc set. Each node $i \in N^K$ corresponds either to the departure of a flight leg $l_i \in L$ on a day $y_i \in Y$, or to its arrival. Each arc represents either a flight leg on a particular day, a deadhead flight on a particular day, or a feasible connection between two flight legs, two deadhead flights or one of each. Connection arcs are defined in the same way as for the aircraft networks. Let also B represent the set of crewbases. For each crewbase $b \in B$, let s_b^K and q_b^K be source and sink nodes that represent, respectively, the beginning and the end of a pairing. One then defines a network $G_b^K = (N_b^K, A_b^K)$ where $N_b^K = N^K \cup \{s_b^K, q_b^K\}$. In addition to the arcs from A^K , set A_b^K contains arcs linking the source node s_b^K to nodes representing the departure of legs that can be covered directly from crewbase b as well as arcs linking nodes that represent the arrival of legs that end at crewbase b to the sink node q_b^K .

Unlike aircraft routes, crew pairings need not be linked together into cycles since the problem of making individual monthly schedules is dealt with separately.

4.1.3 Model

Recall that given two flight legs $l_i, l_j \in L$, the connection between these legs is said to be short if it is feasible but the difference between the departure time of leg l_j and the arrival time of leg l_i is smaller than the minimum sit time for crews. In this case, legs l_i and l_j can be covered, in sequence, by the same crew only if both legs are also covered by the same aircraft. Otherwise, insufficient time is available for the crew to make the connection. Let S be the set of pairs of legs for which the connection between them is short. Given again two flight legs $l_i, l_j \in L$, the connection between these legs is said to be restricted if it is a feasible connection and the difference between the departure time of leg l_j and the arrival time of leg l_i is larger than the minimum sit time for crews but smaller than another given threshold that corresponds to an *ideal* sit time. If a connection is restricted, both legs can be covered in sequence by the same crew, but a penalty is imposed if the legs are not covered in sequence by the same aircraft. Let R be the set of pairs of legs for which the connection between them is restricted.

For every maintenance station $m \in M$ and for every subperiod $t \in T$, let Ω^{mt} be the set of feasible paths from the source node s_{mt}^F to a sink node in G_{mt}^F . Similarly, for every crewbase $b \in B$ let Ω^b denote the set of feasible paths from node s_b^K to node q_b^K in the network G_b^K . A path represents an aircraft route between two maintenance checks or a complete pairing. Each aircraft route must respect a limit on total flight time in addition to the d -day maintenance requirements. Each duty period in a pairing must respect daily limits on total work time, total flight time and total number of landings. In addition, the number of duty periods in a pairing must not exceed a certain limit which is typically four or five. The number of days in a pairing might

differ from the number of duties since the rest period between two consecutive duties does not necessarily take place at night. These path feasibility constraints are modeled through the use of resources and can be handled directly by dynamic programming in a column generation framework (see, e.g., DESAULNIERS *et al.* (1998b)).

For every aircraft path $\omega \in \Omega^{mt}$, define binary constants a_ω^i and b_ω^l that take value 1 if node $i \in N_{mt}^F$ and leg $l \in L$ belong to this path, respectively. Define also binary constants n_ω^{ij} equal to 1 if legs l_i and l_j are performed in sequence in path ω . Let f_ω be the number of aircraft required to cover path ω and s_ω be the number of short connections used by path ω . In a daily problem solved on an acyclic network spanning more than one day, f_ω may take a value greater than one since every leg has to be covered every day. Let also c_ω be the cost of sending one unit of flow along path ω , and let θ_ω be a binary variable that represents the flow on this path. Corresponding parameters a_ω^i , b_ω^l , n_ω^{ij} and c_ω as well as the number of duties in path ω , u_ω , and binary variables χ_ω are defined for every crew path $\omega \in \Omega^b$. For every restricted connection between a leg l_i and a leg l_j , let R_{ij} be a binary variable equal to 1 if the connection is used by a crew without being used by an aircraft, and let p_{ij} be the penalty cost associated with this connection. The penalty cost is equal to a certain number of monetary units for every unit of time below the ideal waiting time. This value can be fixed after a sensibility analysis is performed or calculated with data on hourly crew wages, the probability of missing a connection and other relevant information. Finally, constants ζ^F , ζ^S and ζ^D represent the number of available aircraft, the total number of allowed short connections and a limit on the total number of duties in all crew pairings, respectively. Table 4.1 provides a summary of the notation used in the formulation.

The generalized model for integrated aircraft routing and crew scheduling (GIM) can be stated as follows:

$$\text{Minimize} \quad \sum_{b \in B} \sum_{\omega \in \Omega^b} c_\omega \chi_\omega + \sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} c_\omega \theta_\omega + \sum_{(l_i, l_j) \in R} p_{ij} R_{ij} \quad (4.1)$$

subject to

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} b_\omega^l \theta_\omega = 1 \quad (l \in L) \quad (4.2)$$

$$\sum_{b \in B} \sum_{\omega \in \Omega^b} b_\omega^l \chi_\omega = 1 \quad (l \in L) \quad (4.3)$$

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} f_\omega \theta_\omega \leq \zeta^F \quad (4.4)$$

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} s_\omega \theta_\omega \leq \zeta^S \quad (4.5)$$

$$\sum_{b \in B} \sum_{\omega \in \Omega^b} u_\omega \chi_\omega \leq \zeta^D \quad (4.6)$$

$$\sum_{\omega \in \Omega^{mt}} a_\omega^{q_{mt}^F} \theta_\omega - \sum_{\omega \in \Omega^{mt}} a_\omega^{s_{m,t+1+h/p}^F} \theta_\omega = 0 \quad (m \in M; t \in T) \quad (4.7)$$

$$\sum_{b \in B} \sum_{\omega \in \Omega^b} n_\omega^{ij} \chi_\omega - \sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} n_\omega^{ij} \theta_\omega \leq 0 \quad ((l_i, l_j) \in S) \quad (4.8)$$

$$\sum_{b \in B} \sum_{\omega \in \Omega^b} n_\omega^{ij} \chi_\omega - \sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} n_\omega^{ij} \theta_\omega - R_{ij} \leq 0 \quad ((l_i, l_j) \in R) \quad (4.9)$$

$$R_{ij} \geq 0 \quad ((l_i, l_j) \in R) \quad (4.10)$$

$$\theta_\omega \in \{0, 1\} \quad (m \in M; t \in T; \omega \in \Omega^{mt}) \quad (4.11)$$

$$\chi_\omega \in \{0, 1\} \quad (b \in B; \omega \in \Omega^b). \quad (4.12)$$

The objective function (4.1) minimizes the sum of all crew scheduling, aircraft routing and penalty costs. For the computational experiments presented in Section 4.3, no costs are associated with aircraft routings and we use an approximate crew cost

function that includes waiting costs and fixed costs for deadheads. Details on the cost function are given in Section 4.3.1. Constraints (4.2) and (4.3) ensure that each leg is covered by exactly one aircraft and one crew. Constraint (4.4) imposes a limit on the number of available aircraft. Constraint (4.5) offers the possibility of limiting the total number of short connections used so as to increase the robustness of the produced solutions, where ζ^S can be arbitrarily fixed by the airline. Constraint (4.6) limits the total number of duties worked. This constraint is used to control the duration of each duty in order to reflect the minimum paid flying time, which is not included in our approximate crew cost function. By restricting the number of duties worked, we increase their duration and make unattractive short duties which would incur penalty charges for the airline (the controlling parameter, which in turn modifies ζ^D , is a minimum average duty time for all pairings). Constraints (4.7) make the aircraft routes cyclical by linking paths that start and end at the same maintenance station while ensuring that there is enough time for maintenance between the end and the beginning of the two paths. Finally, constraints (4.8) guarantee that a crew does not change aircraft when the connection time is too short and constraints (4.9) impose a penalty whenever a crew changes aircraft on a restricted connection. The model presented here generalizes upon that introduced by CORDEAU *et al.* (2001) and which is comprised of (4.2), (4.3), (4.8), (4.11) and (4.12). Constraints (4.4) and (4.7) were added to deal with the cyclical aspect of the daily problem while constraint (4.6) helps to compensate for the approximate crew cost function. Finally, constraints (4.5), (4.9) and (4.10) increase the robustness of the model. The basic integrated model (IM) is formulated as the GIM, except that it excludes constraints (4.5), (4.9) and (4.10) as well as variables R_{ij} .

Table 4.1 Summary of notation

B	set of crewbases
M	set of maintenance stations
R	set of pairs of legs for which the connection between them is restricted
S	set of pairs of legs for which the connection between them is short
T	set of possible starting times for aircraft paths
Ω^{mt}	set of feasible paths from the source node s_{mt}^F to a sink node in G_{mt}^F
Ω^b	set of feasible paths from the source node s_b^K to the sink node q_b^K in G_b^K
χ_ω	binary variable that represents the flow on crew path ω
θ_ω	binary variable that represents the flow on aircraft path ω
R_{ij}	binary variable equal to 1 if connection $(l_i, l_j) \in R$ is used by a crew but not by an aircraft
a_ω^i	equal to 1 if node i belongs to path ω
b_ω^l	equal to 1 if leg l belongs to path ω
c_ω	the cost of sending one unit of flow along path ω
f_ω	the number of aircraft required to cover path ω
h	the time required to perform a regular maintenance check
n_ω^{ij}	equal to 1 if legs i and j are performed in sequence in path ω
p	subperiod length
p_{ij}	penalty cost associated with connection $(l_i, l_j) \in R$
q_{mt}^F	sink node for aircraft paths ending at station m before or at o_t and after e_t
q_b^K	sink node for crew paths ending at crewbase b
s_{mt}^F	source node for aircraft paths beginning at station m between o_t and e_t
s_b^K	source node for crew paths beginning at crewbase b
s_ω	the number of short connections used by path ω
ζ^F	the number of available aircraft
ζ^S	the number of short connections allowed
ζ^D	the total number of duties allowed in all crew pairings

4.2 Solution Methodology

When the number of flight legs and the number of short connections each exceed a few hundreds, CORDEAU *et al.* (2001) have shown that solving the integrated model without decomposition becomes difficult. In computational experiments, the authors have observed that computation times seem to grow quadratically with the number of constraints. Using Benders decomposition, their formulation can, however, be decomposed so as to obtain a pair of problems that are easier to solve. Since the generalized model (GIM) introduced in this paper is harder to solve due to a

significantly larger number of constraints, we introduce several improvements to the solution approach proposed by CORDEAU *et al.* (2001). We first propose to use the crew pairing problem as the Benders master problem instead of the aircraft routing problem. We then explain how to improve the generation of Benders cuts so as to reduce the number of iterations needed to reach optimality.

4.2.1 Benders decomposition using the crew pairing problem as a master problem

The generalized integrated model (4.1)-(4.12) includes both crew pairing and aircraft routing path variables. Benders decomposition (see, e.g., BENDERS (1962); NEMHAUSER and WOLSEY (1988)) can be used to reformulate the problem so as to reduce the number of variables at the expense of an increase in the number of constraints. The additional constraints can, however, be generated dynamically only when they are violated. In most applications, only a very small subset of constraints need to be generated for an optimal solution to be identified.

Let χ be the set of solutions satisfying the crew constraints (4.3), (4.6) and (4.12). For given values $\bar{\chi}_\omega$ ($b \in B; \omega \in \Omega^b$) $\in \chi$, the MIP relaxation of model (4.1)-(4.12) reduces to the following *primal subproblem* involving only aircraft variables :

$$\text{Minimize} \quad \sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} c_\omega \theta_\omega + \sum_{(l_i, l_j) \in R} p_{ij} R_{ij} \quad (4.13)$$

subject to

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} b_{\omega}^l \theta_{\omega} = 1 \quad (l \in L) \quad (4.14)$$

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} f_{\omega} \theta_{\omega} \leq \zeta^F \quad (4.15)$$

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} s_{\omega} \theta_{\omega} \leq \zeta^S \quad (4.16)$$

$$\sum_{\omega \in \Omega^{mt}} a_{\omega}^{q_{mt}^F} \theta_{\omega} - \sum_{\omega \in \Omega^{mt}} a_{\omega}^{s_{m,t+1+h/p}^F} \theta_{\omega} = 0 \quad (m \in M; t \in T) \quad (4.17)$$

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} n_{\omega}^{ij} \theta_{\omega} \geq \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \bar{\chi}_{\omega} \quad ((l_i, l_j) \in S) \quad (4.18)$$

$$\sum_{m \in M} \sum_{t \in T} \sum_{\omega \in \Omega^{mt}} n_{\omega}^{ij} \theta_{\omega} + R_{ij} \geq \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \bar{\chi}_{\omega} \quad ((l_i, l_j) \in R) \quad (4.19)$$

$$R_{ij} \geq 0 \quad ((l_i, l_j) \in R) \quad (4.20)$$

$$\theta_{\omega} \geq 0 \quad (m \in M; t \in T; \omega \in \Omega^{mt}). \quad (4.21)$$

Observe that upper bounds on the θ_{ω} variables are not needed because of constraints (4.14). Let $\alpha = (\alpha_l | l \in L)$, $\beta \leq 0$, $\gamma \leq 0$, $\rho = (\rho_{mt} | m \in M, t \in T)$, $\mu = (\mu_{ij} \geq 0 | (l_i, l_j) \in S)$, $\pi = (\pi_{ij} \geq 0 | (l_i, l_j) \in R)$, be the dual variables associated with constraints (4.14)-(4.19), respectively. The dual of (4.13)-(4.21) is the following *dual subproblem*:

$$\begin{aligned} \text{Maximize } & \sum_{l \in L} \alpha_l + \zeta^F \beta + \zeta^S \gamma + \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \bar{\chi}_{\omega} \mu_{ij} + \\ & \sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \bar{\chi}_{\omega} \pi_{ij} \end{aligned} \quad (4.22)$$

subject to

$$\sum_{l \in L} b_{\omega}^l \alpha_l + f_{\omega} \beta + s_{\omega} \gamma + \rho_{mt} - \rho_{m,t+1+h/p} + \sum_{(l_i, l_j) \in S} n_{\omega}^{ij} \mu_{ij} + \sum_{(l_i, l_j) \in R} n_{\omega}^{ij} \pi_{ij} \leq c_{\omega} \quad (m \in M; t \in T; \omega \in \Omega^{mt}) \quad (4.23)$$

$$\pi_{ij} \leq p_{ij} \quad ((l_i, l_j) \in R) \quad (4.24)$$

$$\beta, \gamma \leq 0 \quad (4.25)$$

$$\mu_{ij} \geq 0 \quad ((l_i, l_j) \in S) \quad (4.26)$$

$$\pi_{ij} \geq 0 \quad ((l_i, l_j) \in R). \quad (4.27)$$

Assuming that $c_{\omega} \geq 0$ for all $m \in M$, $t \in T$ and $\omega \in \Omega^{mt}$ and that $p_{ij} \geq 0$ for all $(l_i, l_j) \in R$, the dual subproblem is always feasible since the null vector $\mathbf{0}$ satisfies constraints (4.23)-(4.27). Hence, the primal subproblem is either infeasible or feasible and bounded. Let Δ denote the polyhedron defined by constraints (4.23)-(4.27), and let P_{Δ} and R_{Δ} be the sets of extreme points and extreme rays of Δ , respectively.

One can notice that only the objective function of model (4.22)-(4.27) contains crew information. This implies that the feasible region of the dual subproblem is independent of the crew pairing solution. Introducing the additional free variable z_0 , the MIP relaxation of model (4.1)-(4.12) can thus be reformulated as the following *Benders master problem*:

$$\text{Minimize} \quad \sum_{b \in B} \sum_{\omega \in \Omega^b} c_{\omega} \chi_{\omega} + z_0 \quad (4.28)$$

subject to

$$\begin{aligned}
z_0 - \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \mu_{ij} - \\
\sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \pi_{ij} \geq \sum_{l \in L} \alpha_l + \zeta^F \beta + \zeta^S \gamma \quad ((\alpha, \beta, \gamma, \mu, \pi) \in P_{\Delta}) \quad (4.29) \\
- \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \mu_{ij} - \\
\sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \pi_{ij} \geq \sum_{l \in L} \alpha_l + \zeta^F \beta + \zeta^S \gamma \quad ((\alpha, \beta, \gamma, \mu, \pi) \in R_{\Delta})
\end{aligned} \tag{4.30}$$

$$\sum_{b \in B} \sum_{\omega \in \Omega^b} b_{\omega}^l \chi_{\omega} = 1 \quad (l \in L) \tag{4.31}$$

$$\sum_{b \in B} \sum_{\omega \in \Omega^b} u_{\omega} \chi_{\omega} \leq \zeta^D \tag{4.32}$$

$$\chi_{\omega} \in \{0, 1\} \quad (b \in B; \omega \in \Omega^b). \tag{4.33}$$

Feasibility constraints (4.30) ensure that the values given to the crew variables χ_{ω} ($b \in B; \omega \in \Omega^b$) lead to a bounded dual subproblem. When bounded, the purpose of the dual subproblem is to cost the restricted connection penalties for a specific set of short and restricted connections. The value of z_0 is thus restricted to be larger than or equal to the optimal value of the dual subproblem by optimality constraints (4.29).

In general, model (4.28)-(4.33) contains more constraints than the MIP relaxation of (4.1)-(4.12) but most optimality and feasibility constraints are inactive in an optimal solution. Hence, these constraints need not be enumerated exhaustively but can instead be generated dynamically by iterating between a relaxed master problem and the subproblem. The relaxed master problem contains constraints (4.31)-(4.33) as well as subsets of Benders cuts (4.29) and (4.30). The optimal solution of the relaxed

Benders master problem is used to set up constraints (4.18) and (4.19) in the primal subproblem at every iteration. If the primal subproblem is feasible, the value of the dual variables associated with constraints (4.14)-(4.19) determine an extreme point of P_{Δ} . Otherwise, an extreme ray of R_{Δ} violating one of the constraints (4.30) is identified. Hence, exactly one constraint is added to the relaxed Benders master problem at each iteration and the process continues until its optimal solution yields a feasible primal subproblem whose cost is equal to the value of z_0 .

Disregarding the penalties attached to the restricted connections, the aircraft routing problem reduces to a feasibility problem. Most of the costs in the integrated model are thus associated with the crew pairings. The benefit of reversing the natural solution sequence (the order usually chosen in a sequential approach) in the Benders decomposition method is that the subproblem is then mostly transferring feasibility information to the master problem and very little optimality (or cost) information. As a result, one may expect the number of generated cuts to decrease with respect to a decomposition in which the aircraft routing problem acts as the master problem.

4.2.1.1 Three-phase algorithm

CORDEAU *et al.* (2001) have described a heuristic solution method in which both the aircraft routing master problem and the crew pairing subproblem are solved by column generation. The method works in three phases. In Phase I, all integrality requirements are relaxed and the LP relaxation of the integrated model is solved to optimality by Benders decomposition and column generation. Retaining all generated cuts, Phase II reintroduces integrality constraints on the master problem variables, and solves the resulting mixed-integer problem by generating additional cuts. In this phase, the integer master problem must be solved at each iteration of the Benders decomposition algorithm. In Phase III, integrality constraints are finally added on

the subproblem variables and the integer subproblem is solved once with the values of the master problem variables being held fixed. Since the subproblem does not have the integrality property, the optimal value of the integer subproblem may exceed that of the linear relaxation, in which case the optimality of the solution to the integrated model is not guaranteed. Because it also includes an unknown integrality gap, we call *maximum optimality gap* the percentage difference between the integer solution of the integrated problem, obtained after Phase III, and the LP relaxation solution computed in Phase I. (To get an optimal integer solution to the integrated model, one would have to embed the three-phase approach in an external branch-and-bound algorithm.) By first solving the LP relaxation of the problem in Phase I, the authors have observed an overall performance increase since very few cuts are generated in Phase II when the relaxed master problem has to be solved at each iteration with integrality constraints.

The solution methodology that we use here is also based on this heuristic three-phase Benders decomposition and column generation approach. The large size of the sets Ω^b in the relaxed Benders master problem (4.28)-(4.33) and of the sets Ω^{mt} in the Benders subproblem (4.13)-(4.21) prohibits a complete enumeration of all aircraft and crew path variables. However, the linear relaxation of each of these two models can be solved with a column generation algorithm in the first phase of the Benders decomposition approach and at each node of the branch-and-bound tree in the second and third phases. At each iteration of the column generation process, a restricted column generation master problem is solved using only a subset of all variables. This problem is obtained by replacing the sets Ω^b or Ω^{mt} by the subsets $\Omega_\tau^b \subseteq \Omega^b$ or $\Omega_\tau^{mt} \subseteq \Omega^{mt}$ of variables available at iteration $\tau = 0, 1, \dots$. New variables for the column generation master problem are then generated at each iteration by solving a resource-constrained shortest-path problem for each network G_b^K ($b \in B$) or G_{mt}^F ($m \in M, t \in T$), (see, e.g., DESAULNIERS *et al.* (1998b); BARNHART *et al.* (1998b)).

Path variables can indeed be identified by sending one unit of flow from the source to the sink of the networks. At each iteration, arc costs are modified to reflect the current values of the dual variables associated with the constraints of the restricted column generation master problem. The column generation process stops when no negative-cost path can be identified in any network.

With the reversed Benders decomposition approach, linking constraints in the primal subproblem are forcing the aircraft to use the short connections used by the crews in the master problem. This could potentially lead to an infeasible integer primal subproblem (Phase III) for the given solution of the master problem even if the original problem is feasible. A step is therefore added after the third phase to verify the feasibility of the integer subproblem and, if needed, go back to the second phase to solve the integer master problem with an additional constraint forbidding the same subset of short connections to be chosen. Going through the second and third phases repeatedly could greatly increase the computing time, but our computational experience has shown that only a small number of these iterations are necessary. In addition, since most of the costs are in the master problem, the maximum optimality gap related to the subproblem is small. The need for an external branch-and-bound algorithm is therefore less important with the reversed Benders decomposition approach.

To compute an optimal integer solution of either the Benders master problem or the subproblem, different branching schemes can be devised. A possible scheme consists of branching directly on the path variables χ_ω and θ_ω . However, because imposing the constraint $\chi_\omega = 0$, for example, is difficult to accomplish when using column generation, one must often be content with finding heuristic solutions by a depth-only search. In this case, a fractional variable χ_ω is selected at each node of the tree and the constraint $\chi_\omega = 1$ is added to the restricted column generation master problem. The improved Benders decomposition approach includes a variant of this

branching scheme. At each node of the branching tree, when one or several variables take fractional values larger than or equal to a certain threshold, λ_1 (close to 1), branching decisions are made simultaneously on all such variables to accelerate the search. If no fractional variable takes a value larger than or equal to λ_1 , a certain number of variables taking fractional values that exceed a second threshold, $\lambda_2 \leq \lambda_1$, are sequentially fixed to 1 and the resulting LP's are solved. Only the newly added node with the most promising bound is kept in the tree and the others are removed. If still no decision can be made at that point (because all fractional variables take a value smaller than λ_2), branching is done on arcs or follow-ons in the shortest-path networks. This branching strategy is modified toward the end of the tree where complete enumeration is performed. The process is repeated until all variables satisfy integrality requirements or the column generation restricted master problem becomes infeasible. This branch-and-bound scheme helped in reducing variance in both solution quality and computing time.

4.2.2 Identifying strong cuts

Whenever the primal subproblem (4.13)-(4.21) is degenerate, there may exist more than one optimal solution to the dual subproblem. Although any of these points leads to a valid optimality cut, some may yield stronger cuts than others. The cut generated from the extreme point $(\alpha^1, \beta^1, \gamma^1, \mu^1, \pi^1)$ dominates the cut generated from the extreme point $(\alpha^2, \beta^2, \gamma^2, \mu^2, \pi^2)$ if and only if

$$\begin{aligned} \sum_{l \in L} \alpha_l^1 + \zeta^F \beta^1 + \zeta^S \gamma^1 + \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \mu_{ij}^1 + \sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \pi_{ij}^1 &\geq \\ \sum_{l \in L} \alpha_l^2 + \zeta^F \beta^2 + \zeta^S \gamma^2 + \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \mu_{ij}^2 + \sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega} \pi_{ij}^2 & \end{aligned}$$

for all χ_ω ($b \in B; \omega \in \Omega^b$) $\in \chi$ with strict inequality for at least one point. A cut is said to be Pareto-optimal if no other cut dominates it (MAGNANTI and WONG, 1981).

Let χ^{LP} be the polyhedron defined by (4.3), (4.6) and constraints $\chi_\omega \geq 0$ ($b \in B; \omega \in \Omega^b$), and let $ri(\chi^{LP})$ denote the relative interior of χ^{LP} . For a given solution $\bar{\chi} \in \chi^{LP}$ for which the primal subproblem is feasible, let $v(\bar{\chi})$ denote the optimal value of the subproblem. To identify an optimal solution to the dual subproblem that yields a Pareto-optimal cut, one must solve the following dual auxiliary problem, where χ^0 is a point chosen in $ri(\chi^{LP})$:

$$\begin{aligned} \text{Maximize } & \sum_{l \in L} \alpha_l + \zeta^F \beta + \zeta^S \gamma + \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega}^0 \mu_{ij} + \\ & \sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \chi_{\omega}^0 \pi_{ij} \end{aligned} \quad (4.34)$$

subject to

$$\begin{aligned} & \sum_{l \in L} \alpha_l + \zeta^F \beta + \zeta^S \gamma + \sum_{(l_i, l_j) \in S} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \bar{\chi}_{\omega} \mu_{ij} + \\ & + \sum_{(l_i, l_j) \in R} \sum_{b \in B} \sum_{\omega \in \Omega^b} n_{\omega}^{ij} \bar{\chi}_{\omega} \pi_{ij} = v(\bar{\chi}) \end{aligned} \quad (4.35)$$

$$(\alpha, \beta, \gamma, \mu, \pi) \in \Delta. \quad (4.36)$$

Besides including the same dual subproblem constraints (4.36), the dual auxiliary problem contains an additional constraint (4.35) to ensure that one will choose an extreme point from the set of optimal solutions to the original dual subproblem (i.e., with a value equal to the original optimal value $v(\bar{\chi})$). The objective function of the auxiliary problem (4.34) compares all possible cuts at a point χ^0 in the feasible region of the master problem. It therefore tries to maximize the strength of the

cut to be added to the Benders master problem. Instead of solving model (4.34)-(4.36), one can solve the primal form. This approach is convenient in terms of ease of implementation and computational efficiency since the same basic representation can then be used to solve both the primal subproblem and the auxiliary problem. Only the right-hand sides of the linking constraints are modified to reflect the objective function (4.34) and an additional variable associated to constraint (4.35) is introduced in the model. Pareto-optimal cuts generated with the auxiliary problem are cuts generated from extreme points. They are thus optimality cuts. To ensure that the primal subproblem (4.13)-(4.21) is feasible for any solution of the relaxed master problem and that no feasibility cuts are therefore necessary, one can add artificial variables with large positive costs in the primal subproblem. Constraints (4.30) are then no longer necessary because the dual subproblem polyhedron (and at the same time the dual auxiliary problem polyhedron) becomes bounded and contains no ray.

Since the master problem is solved by column generation, its set of variables is not fully known. Because an explicit representation of the polyhedron is therefore not available, finding a solution χ^0 in $ri(\chi^{LP})$ appears to be a rather difficult task. Although it might alter the strength of the cut, choosing a solution χ^0 which is not in $ri(\chi^{LP})$ does not prevent the auxiliary problem to generate a valid cut for the master problem since χ^0 only modifies the objective function of the auxiliary problem (4.34)-(4.36). Hence, we can arbitrarily fix the coefficients of the variables μ_{ij} and π_{ij} in (4.34) to generate a valid Benders cuts. Experiments have shown that values close to 0 generate stronger cuts than values close to 1 when the subproblem is the aircraft routing problem. As a rule of thumb, it seems that making the resulting constraint easier to satisfy in the primal form of the auxiliary problem yields a stronger cut. One can also verify that the same logic applies when the subproblem is the crew pairing problem, but since the linking constraints are reversed in that case, values close to 1 give constraints that are easier to satisfy and yield stronger cuts than values close to 0. Another possibility is

to compute a convex linear combination of prior Benders master problem solutions as an estimate of a point in $ri(\chi^{LP})$ at every iteration, but computational experiments have shown that directly fixing the coefficients of μ_{ij} and π_{ij} in (4.34) at a value close to 0 better improves the strength of the Benders cuts. Since all the cuts generated with the auxiliary problem are valid Benders cuts, convergence is still guaranteed.

4.2.3 Step-by-step description of the improved algorithm

Let η represent the iteration number and let P_{Δ}^{η} represent the restricted sets of extreme points of Δ available at iteration η . Let the relaxed master problem be the problem obtained from the master problem (4.28)-(4.33) by replacing P_{Δ} by P_{Δ}^{η} . Artificial variables are added in the primal subproblem so that the dual subproblem and the dual auxiliary problem no longer contain any ray.

Phase I

1. Remove integrality constraints on all variables.
2. Set $\eta := 1$, $P_{\Delta}^1 := \emptyset$ and choose a point χ^0 in $ri(\chi^{LP})$.
3. Solve the relaxed master problem with column generation.
 - (a) If the master problem is infeasible, then the original problem is infeasible, **stop**.
 - (b) Otherwise, let $(\bar{\chi}^{\eta}, z_0^{\eta})$ be an optimal solution.
4. Solve the primal subproblem with column generation, taking $\bar{\chi}^{\eta}$ as an input, and let $\bar{\theta}^{\eta}$ be a primal optimal solution of value $v(\bar{\chi}^{\eta})$. (This problem is always feasible since artificial aircraft variables are included.)

- (a) If $v(\bar{\chi}^\eta) = z_0^\eta$, then $(\bar{\theta}^\eta, \bar{\chi}^\eta)$ is an optimal solution to the problem, **stop this phase**.
5. Solve the auxiliary subproblem with column generation, taking $\bar{\chi}^\eta$, $v(\bar{\chi}^\eta)$ and χ^0 as inputs, and let $(\alpha, \beta, \gamma, \mu, \pi)^\eta$ be a dual optimal solution given as an extreme point. (This problem is always feasible since artificial aircraft variables are included.)
- (a) Set $P_{\Delta}^{\eta+1} := P_{\Delta}^\eta \cup \{(\alpha, \beta, \gamma, \mu, \pi)^\eta\}$ to generate an optimality cut.
- (b) Set $\eta := \eta + 1$ and return to Step 3.

Phase II

1. Reintroduce integrality constraints on the master problem variables (crew variables).
2. Return to Step 3 of Phase I (solve the relaxed master problem by generating new columns at every node of the branch-and-bound tree in Step 3).

Phase III

1. Reintroduce integrality constraints on the subproblem variables (aircraft variables).
2. Solve the primal subproblem with column generation (generating new columns at every node of the branch-and-bound tree), taking $\bar{\chi}^\eta$ as an input.
 - (a) If the subproblem is feasible and no artificial variable is used in the solution, let $\bar{\theta}^\eta$ be a primal optimal solution. Then $(\bar{\theta}^\eta, \bar{\chi}^\eta)$ is a heuristic solution to the original problem, **stop**.

- (b) Otherwise, add a constraint to the relaxed master problem to forbid the set of short connections used by $\bar{\chi}^n$, remove integrality constraints on the subproblem variables and return to Step 2 of Phase II.

The LP solution obtained at the end of Phase I is optimal. However, the MIP solution obtained at the end of Phase II is optimal only if the branch-and-bound tree is fully evaluated, which is not the case with the branching strategy described in Section 4.2.1.1 since not all branching decisions permit backtracking. Phase III IP solution is heuristic no matter which branching strategy is used since the integer subproblem is solved with the values of the master problem variables being held fixed (the three-phase approach is not embedded in an external branch-and-bound algorithm). However, since most of the costs are in the master problem, the maximum optimality gap related to the subproblem is small.

4.3 Computational Experiments

In this section, we present computational experiments that were carried out on instances based on data provided by two major airlines. We first provide a description of these instances, followed by a summary of our computational experiments.

4.3.1 Description of data sets

The test instances come from daily fleet assignment solutions provided by the airlines. Only aircraft types assigned to 150 flights or more were kept, which yielded eight instances. To determine the number of available aircraft of each type, an aircraft routing problem was solved for every instance, with the objective of minimizing the

number of aircraft used under constraints (4.2), (4.7) and (4.11). The value of the right-hand side of constraint (4.6), i.e., ζ^D , was set so that the average duty time in the pairings is greater than or equal to 5.5 hours, a figure which is in line with the results obtained by analyzing real airline crew solutions.

The characteristics of the different instances are summarized in Table 4.2. This table indicates the number of legs, the number of short connections (SC), the number of restricted connections (RC), the number of aircraft and the percentage of linking constraints in the total number of constraints. Here, a connection between two flight legs l_i and l_j is feasible if the difference between the departure time of leg l_j and the arrival time of leg l_i is greater than or equal to a station-specific minimum plan turn time ψ_s that varies between 30 and 45 minutes. However, the connection is said to be short if the connection time exceeds ψ_s by less than 30 minutes. Hence, if ψ_s is equal to 30 minutes, all connections whose duration lies between 30 and 60 minutes are said to be short and the ones that are between 60 and 90 minutes are said to be restricted. One can notice that the difficulty of solving the integrated model as opposed to the individual aircraft routing and crew pairing problems increases with the number of legs since the percentage of linking constraint also increases with this parameter.

Table 4.2 Characteristics of test instances

Instance	Legs	SC	RC	Aircraft	% linking
D95A	196	89	47	41	24%
B757A	184	114	43	60	29%
B767R	151	84	144	38	35%
A320B	258	183	139	80	37%
D9SA	523	502	270	116	41%
B767S	510	370	612	130	48%
D9SB	508	659	467	116	51%
MD80L	707	1183	889	143	59%

In our experiments, pairs of flight legs to be covered by the same aircraft were specified directly by the airlines. Hence, through values need not be considered explicitly in the aircraft routing problem and the only costs that remain in the problem are the

operational costs that are associated with the flight legs. But because all legs must be covered exactly once and all aircraft of a given type are assumed to have equal operating costs, the operational costs are in fact a fixed cost. The aircraft routing problem thus becomes a feasibility problem to ensure that each aircraft is maintained appropriately, except for the possible restricted connection penalties.

For most airlines, the cost of a duty period is proportional to the maximum of two quantities: the total flying time and the minimum guaranteed number of hours per duty period. The cost of a pairing is also roughly equal to the maximum of two quantities: the sum of the costs of the individual duties that constitute the pairing and a fraction of the total duration of the pairing (or *time away from base*). In addition, when an overnight rest takes place in a city that is not the crewbase, additional expenses are incurred for hotel, meals and ground transportation. Finally, *deadhead* costs are incurred when crew members travel as passengers on certain flights. Deadheads are useful to reposition crews to a different city where they are needed to cover a flight leg. They can also be used to ensure that the crew can return to its base at the end of a pairing.

We used an approximate crew cost function which includes waiting costs and fixed costs for deadheads (restricted to the company's own flights). Because each flight leg must be covered by exactly one crew, a large portion of total crew costs is fixed. Hence, the only relevant costs considered in these experiments are those that can be reduced by a better planning of crew pairings. Variable expenses are incurred for connections whose duration exceeds a given threshold because crews must then be credited work time even though they are not actually working. Additional accommodation expenses are also incurred when the rest period between successive duties does not take place at the crewbase. These waiting costs are illustrated by the piecewise linear function represented in Figure 4.1. The first three pieces of the function deal with sits and the last three with rests. A sitting time of 90 minutes is considered ideal, while at

least 60 minutes is required when crews do not follow an aircraft. A waiting time of 60 minutes for a crew that does not follow an aircraft might create some delay and therefore a decreasing penalty is imposed on connections that last between 60 and 90 minutes to make the crew solutions more robust. For our computational experiments, we calculated the level of penalty of these restricted connections, p_{ij} , using data on hourly crew wages, the probability of missing a connection and the average duty length. Penalizing restricted connections increases the need for a model and a solution approach that handle crew and aircraft decisions simultaneously so as to obtain an improved global solution. The third piece of the function represents a compensation given to the crews when the connection time is below ten hours and, hence, is too short to be a rest taken at a hotel. Those unpleasant waits are accounted for by imposing a cost for sitting time in excess of the ideal 90 minutes. When waiting time exceeds ten hours, it is considered a rest and a fixed cost, including hotel and transportation costs, is imposed. Lastly, when a rest exceeds twelve hours, it is deemed too long and waiting time passed this limit is compensated at the same rate as long sits, in addition to the fixed rest cost. When a crew follows an aircraft, the cost function does not comprise the first two pieces and waiting times from 30 to 90 minutes are free of charge.

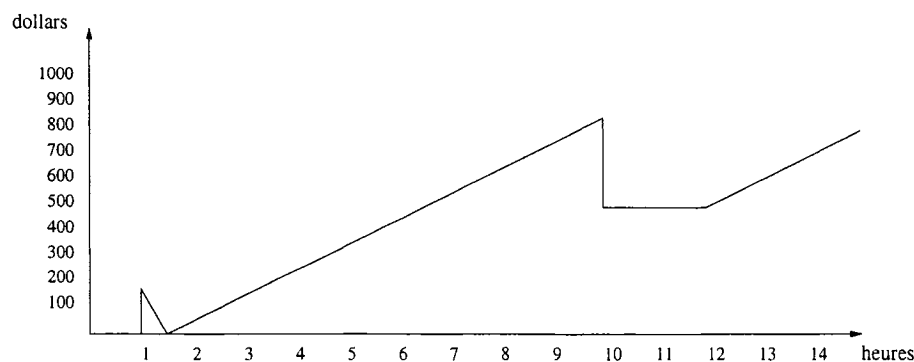


Figure 4.1 Waiting time cost function

4.3.2 Summary of computational experiments

To evaluate the benefits of solving the crew pairing problem in the Benders master problem (*MP Crew*) as opposed to the aircraft routing problem (*MP Aircraft*), we first tried to solve each of the eight instances with both methods. We then included the solution of an auxiliary aircraft routing problem in phases I and II to generate Pareto-optimal cuts (*MP CPO*). The coefficients of μ_{ij} and π_{ij} in (4.34) were first set equal to a linear combination of prior relaxed Benders master problem solutions, but fixing them to a value close to 0 proved to be better. We present here the results obtained by fixing the coefficients of μ_{ij} and π_{ij} equal to 0.10. We tested the three approaches on both the basic integrated model (IM) and the generalized integrated model (GIM). We also tried to generate Pareto-optimal cuts in the basic Benders decomposition approach (*MP Aircraft*) but the time needed to solve the harder auxiliary crew scheduling problem was always greater than or equal to the time saved by having a reduced number of generated cuts. Our algorithms were coded in C++ and use GENCOL¹ for column generation. All experiments were performed on a Pentium 4, 2 GHz computer, using a single processor.

The LP relaxation (Phase I) is not always solved to optimality. Instead, cuts are generated until the relative difference between the lower bound provided by the Benders relaxed master problem and the upper bound provided by the subproblem is less than or equal to a given value. This relative difference was alternately fixed to 1%, 0.1% and 0.01%. Using a larger tolerance obviously speeds up the completion of Phase I, but experience has shown that the quality of the integer solutions found in Phase III is significantly lower in that case. Table 4.3 gives average maximum optimality gaps (over the eight instances) which are computed with respect to the best LP relaxation value known (obtained when solving with a tolerance of 0.01%).

¹GENCOL is an optimization software that was developed at GERAD in Montreal.

Table 4.3 Phase III Solution quality comparisons (% average maximum optimality gap)

Opt.Tol.	IM			GIM		
	MP Aircraft	MP Crew	MP CPO	MP Aircraft 3 inst.	MP Crew 5 inst.	MP CPO
1%	23.99%	1.13%	1.00%	1.60%	1.55%	1.35%
0.1%	1.58%	1.04%	1.09%	1.37%	1.13%	1.27%
0.01%	1.17%	1.01%	0.99%	1.33%	-	1.22%

For the basic integrated model (IM) with MP Aircraft, one can notice that the average solution quality is very poor with a 1% tolerance. The main reason for this large average gap is that, for two instances, Phase II could not find a feasible solution without resorting to the use of artificial variables with large positive costs, which contributed in deteriorating solution quality. There is also a significant difference in average solution quality between a tolerance of 0.1% and one of 0.01%. Choosing the right tolerance level is less of an issue for the methods involving the reversed Benders decomposition approach because the cuts are often strong enough for the smallest tolerance level (even optimality in Phase I) to be attained without requiring it. Solution quality is nevertheless slightly higher with the smallest tolerance.

Not all instances of the harder generalized integrated model (GIM) could be solved with the MP Aircraft approach. In fact, only three instances could be solved in reasonable computing time with a 0.01% tolerance level and five instances could be solved with 0.1% and 1% tolerance levels. When solving these five instances, one can notice that solution quality is better with a tolerance of 0.1%. Even for GIM, the tolerance levels do not have much impact on solution quality when the reversed Benders decomposition approach is used. In further computational experiments, the Benders optimality tolerance was set to 0.01%, except for GIM with MP Aircraft, where it was fixed at 0.1% to be able to solve more instances.

Table 4.4 Complete results: basic integrated model (IM)[†]

	D95A	B757A	B767R	A320B	D9SA	B767S	D9SB	MD80L	Avg
<i>Cost LP</i>	30,01	37,30	38,32	66,93	80,48	114,55	67,60	91,30	
MP Aircraft									
<i>CPU Ph.I</i>	2.01	1.50	0.35	8.04	20.34	27.22	213.32	1467.79	
<i>CPU Ph.II</i>	0.10	0.74	0.48	6.23	1.58	314.80	128.65	413.05	
<i>CPU Ph.III</i>	1.03	0.16	0.16	2.50	0.47	34.10	1.31	6.25	
<i>CPU Total</i>	3.14	2.40	0.99	16.77	22.39	376.12	343.28	1887.09	
<i>Cuts Ph.I</i>	30	18	5	80	25	62	71	260	
<i>Cuts Ph.II</i>	0	1	0	2	0	28	0	11	
<i>Cost IP</i>	30,61	37,60	38,69	68,07	80,48	116,66	68,23	92,47	
<i>Gap</i>	1.95%	0.79%	0.95%	1.68%	0.00%	1.81%	0.92%	1.27%	1.17%
MP Crew									
<i>CPU Ph.I</i>	0.43	0.48	0.20	2.41	9.64	3.98	13.11	26.57	
<i>CPU Ph.II</i>	2.05	2.66	0.22	15.71	1.47	97.96	1.37	157.66	
<i>CPU Ph.III</i>	0.03	0.11	0.55	0.02	1.28	4.36	7.67	16.16	
<i>CPU Total</i>	2.51	3.25	0.97	18.14	12.39	106.30	22.15	200.39	
<i>Cuts Ph.I</i>	4	3	1	11	6	4	12	6	
<i>Cuts Ph.II</i>	0	2	0	1	0	2	0	1	
<i>SPIP Cuts</i>	0	0	0	0	0	0	0	0	
<i>Cost IP</i>	30,49	37,81	38,70	67,74	80,48	116,59	68,20	91,69	
<i>Gap</i>	1.55%	1.34%	0.97%	1.20%	0.00%	1.75%	0.87%	0.42%	1.01%
<i>CPU vs 1</i>	1.25	0.74	1.02	0.92	1.81	3.54	15.50	9.42	4.27
MP CPO									
<i>CPU Ph.I</i>	0.44	0.35	0.28	1.31	5.85	3.34	9.69	16.22	
<i>CPU Ph.II</i>	0.58	0.55	0.10	0.81	0.39	46.21	1.76	14.49	
<i>CPU Ph.III</i>	0.04	0.09	0.72	0.14	1.99	3.63	6.37	25.29	
<i>CPU Total</i>	1.06	0.99	1.10	2.26	8.23	53.18	17.82	56.00	
<i>Cuts Ph.I</i>	2	1	1	4	2	1	3	1	
<i>Cuts Ph.II</i>	0	1	0	0	0	0	1	0	
<i>SPIP Cuts</i>	0	0	0	0	0	0	0	0	
<i>Cost IP</i>	30,41	37,60	38,70	67,92	80,48	116,42	68,20	92,18	
<i>Gap</i>	1.30%	0.79%	0.97%	1.46%	0.00%	1.60%	0.87%	0.95%	0.99%
<i>CPU vs 1</i>	2.96	2.42	0.90	7.42	2.72	7.07	19.26	33.70	9.56
<i>CPU vs 2</i>	2.37	3.28	0.88	8.03	1.51	2.00	1.24	3.58	2.86

[†] All CPU times are in minutes.

Tables 4.4 and 4.5 present a comparison of the CPU time and computational effort needed to solve all three phases with the different approaches on both the basic and the generalized integrated models. $Cost \widehat{LP}$ indicates the cost of the best solution found at the end of Phase I. This constitutes an upper bound on the true LP relaxation value of the integrated problem because the LP relaxation is not necessarily solved to optimality, but with a tolerance of 0.01% or 0.1% (optimality was nevertheless reached for all eight instances on IM and for three instances on GIM). We indicate, for the three approaches and all eight instances, the time spent in each of the three phases as well as the number of cuts generated in the first two phases and the number of forbidden short connection subset cuts added in phase II to get a feasible integer subproblem (*SPIP Cuts*). $Cost IP$ indicates the cost of the integer integrated solution at the end of Phase III. The percentage gap reported is computed with respect to $Cost \widehat{LP}$. Finally, the CPU time efficiency of the methods are compared. For example, *CPU vs 1* reported for the second method is the total CPU time of the first method divided by the total CPU time of the second one.

When considering Phase I CPU times, one observes that the approaches using the reversed Benders decomposition method (MP Crew and MP CPO) always produced LP solutions faster than the original Benders decomposition method (MP Aircraft) and that the difference in performance increases with problem size, especially for the generalized model. The comparison of the time spent finding an integer solution can be misleading because of the heuristic branch-and-bound method, but for every instance, except three of the smaller ones, both the second and third methods produced integer solutions faster than the MP Aircraft method. For the three larger instances, the MP aircraft three-phase approach could not converge in reasonable computing time, even within a 1% optimality tolerance. Phase I could not even converge for two of those problems.

Reversing the natural solution sequence can decrease the average total CPU time by a factor of 4.27 for IM and by a factor of 7.34 for GIM. This last improvement is in fact underestimating the real ratio of decrease since the MP Aircraft method was sometimes stopped before getting a solution. The difference in performance can be attributed, mainly, to the reduced number of cuts generated in Phase I. When Pareto-optimal cuts are generated, the average total CPU time is decreased, on average, by another factor of 2.86 for IM and 3.45 for GIM. Globally, one can see that the improved Benders decomposition approach decreased the average CPU time by a factor of 9.56 for IM (with a decrease of a factor 33 for the larger instance) and by a factor of at least 21.37 for GIM. One can finally notice that these performance improvements do not come at the price of lower solution quality. Tables 4.4 and 4.5 show that the average maximum optimality gaps are actually lower when using the improved Benders decomposition approach.

4.3.3 Comparisons with the extended crew pairing approach of Cohn and Barnhart (2003)

COHN and BARNHART (2003) developed a method to solve the basic integrated model where aircraft solutions are generated first and included as variables in an extended crew pairing problem (ECP). The authors showed that only solutions including unique and maximal short connection sets (UM) needed to be generated to obtain an optimal solution. To obtain unique and maximal sets, one needs to solve successive aircraft maintenance routing problems with the objective of maximizing the number of short connections used. After each UM is generated, an additional constraint has to be included in the aircraft routing problem to prevent the next UM generated to be a subset of an already generated UM. The authors report some computational results on two small instances both containing approximately 125 legs. Not all UM columns were

Table 4.5 Complete results: generalized integrated model (GIM)[†]

	D95A	B757A	B767R	A320B	D9SA	B767S	D9SB	MD80L	Avg
<i>Cost LP</i>	30,81	37,68	39,28	67,69	82,35	117,08	68,56	92,62	
MP Aircraft									
<i>CPU Ph.I</i>	2.83	2.00	11.33	7.34	87.74	487.53	>40h	>40h	
<i>CPU Ph.II</i>	0.39	14.34	7.88	533.84	21.72	>32h			
<i>CPU Ph.III</i>	0.30	0.12	0.16	0.67	0.70				
<i>CPU Total</i>	3.52	16.46	19.37	541.85	110.16	>40h	>40h	>40h	
<i>Cuts Ph.I</i>	32	24	65	52	82	301			
<i>Cuts Ph.II</i>	0	10	1	68	0				
<i>Cost IP</i>	31,20	38,47	39,60	68,64	82,46				
<i>Gap</i>	1.25%	2.06%	0.81%	1.39%	0.14%				1.13%
MP Crew									
<i>CPU Ph.I</i>	0.89	0.94	0.29	4.80	51.20	16.73	79.92	221.08	
<i>CPU Ph.II</i>	0.37	16.56	0.16	355.31	2.22	32.66	33.83	56.13	
<i>CPU Ph.III</i>	0.01	1.35	4.66	0.70	4.60	48.97	4.15	482.24	
<i>CPU Total</i>	1.27	18.85	5.11	360.81	58.02	98.36	117.90	759.45	
<i>Cuts Ph.I</i>	11	6	1	17	50	14	28	30	
<i>Cuts Ph.II</i>	0	34	0	29	0	0	5	16	
<i>SPIP Cuts</i>	0	4	0	0	0	0	0	1	
<i>Cost IP</i>	31,20	38,49	40,09	68,29	82,54	117,71	69,38	94,07	
<i>Gap</i>	1.25%	2.11%	2.01%	0.88%	0.24%	0.54%	1.19%	1.54%	1.22%
<i>CPU vs 1</i>	2.77	0.87	3.79	1.50	1.90	>24.40	>20.36	>3.16	>7.34
MP CPO									
<i>CPU Ph.I</i>	0.93	0.93	0.76	2.64	31.50	12.94	89.38	277.61	
<i>CPU Ph.II</i>	0.23	6.70	0.13	22.71	27.94	6.11	16.58	15.44	
<i>CPU Ph.III</i>	0.05	0.80	3.69	0.12	0.48	2.99	3.99	5.94	
<i>CPU Total</i>	1.21	8.43	4.58	25.47	59.92	22.04	109.95	298.99	
<i>Cuts Ph.I</i>	6	4	4	7	14	4	14	27	
<i>Cuts Ph.II</i>	0	3	0	1	5	0	0	0	
<i>SPIP Cuts</i>	0	5	0	0	2	0	1	0	
<i>Cost IP</i>	31,26	38,44	39,90	68,12	82,53	118,44	69,37	93,18	
<i>Gap</i>	1.45%	1.98%	1.55%	0.64%	0.22%	1.15%	1.17%	0.61%	1.09%
<i>CPU vs 1</i>	2.91	1.95	4.23	21.27	1.84	>108.89	>21.83	>8.03	>21.37
<i>CPU vs 2</i>	1.05	2.24	1.12	14.17	0.97	4.46	1.07	2.54	3.45

[†] All CPU times are in minutes.

generated, but the authors nevertheless obtained solutions with a maximal optimality gap of 0% for the first instance and 1.7% for the second one (ECP solutions compared to solutions where all short connections are included). In fact, ten UM columns were generated sequentially in a first step. Then, six other UM columns were added for the first instance and 10 for the second one to ensure that every possible short connection was included in at least one UM. No details are given on the computation time required to solve the two small instances. Their conclusions were that even with a small subset of UM columns included, the ECP could find good quality solutions. We considered the method proposed by the authors as promising and we therefore applied it out on our test instances.

The same column generation methods developed to solve the aircraft routing problem and the crew scheduling problem in our Benders decomposition approach were used, with the same parameters, to solve either the UM generation problem or the ECP. All the cumbersome constraints related to the increased robustness (constraints (4.5), (4.9) and (4.10)), were taken out since the UM generation problem cannot handle the restricted connection penalties.

The ECP is solved quickly with our column generation algorithm. The time-consuming portion is the UM generation problem. This may be surprising at first sight since the aircraft routing problem is usually considered to be the easiest of the planning problems. However, the UM generation problem is not a simple aircraft routing problem because the objective is to maximize the number of short connections used (as opposed to a feasibility problem) and some complicating constraints are added to get unique short connection sets. We first tried to use an exact branching scheme since the added uniqueness cuts are not valid when generated from a heuristic solution. Even though the first integer UM is very long to find, we can substantially increase the speed of the exact branch-and-bound method, after the first iteration, as a good upper bound is known at that point. Despite that, too many successive aircraft

routing problems have to be solved and after more than 200 hours of computing time, not one of the eight instances was completely solved. We then tried to stop the algorithm after a certain number of UM columns were found and at the same time allow a larger optimality gap, but after 1000 minutes, even generating one column was impossible for the two larger instances and this was not considered reasonable since the improved Benders decomposition method solves the basic integrated problem in less time than that. We finally tried the same heuristic branch-and-bound method used in the Benders decomposition approach. The impact of a heuristic approach (generating non-maximal short connection subsets) is a premature stopping of the algorithm, limiting the number of UM columns that can be generated. We tried to generate 10 UM columns, but only got an average of six columns over the eight instances. Most of the time, only one short connection is switched with another from one UM generation iteration to the next. This means that many short connections are not included at all in the ECP after the heuristic UM generation. In fact, the average percentage of short connections included on the total number of possible SC is 37%. In that case, there is no guarantee on the quality of the overall solution, but it also means that trying to include every short connection in at least one UM cannot be done in a reasonable computing time since a possibly very large number of additional iterations would be needed on top of an already long process for the larger instances.

Although they did not use this approach in their computational experiments, Cohn and Barnhart suggested to initialize the ECP with a small number of UM columns and to generate additional columns of negative reduced-cost dynamically while solving the problem. We think this could be beneficial, but only to a certain extent. Indeed, this approach requires that an aircraft routing problem be solved at each node of the branch-and-bound tree to determine whether UM columns of negative reduced-cost exist. However, our experience shows that finding a very small subset of heuristic UM

columns is already very time-consuming for the larger instances. We therefore solved the ECP with a heuristic subset of UM columns and compared the results (*UM + ECP*) with the ones obtained with the improved Benders decomposition method (*MP CPO*) for the basic integrated model (IM). The results are given in Table 4.6 where the total CPU time, the number of short connections used in the solution, the cost of the integer solution found and the IP maximum optimality gap (*IP MOG*, computed with respect to the 0.01% tolerance LP relaxation value) are given for both methods and all instances. Since the 0.01% tolerance LP relaxation values obtained with the improved Benders decomposition method for IM are in fact the LP relaxation optimal values, we have computed the true percentage LP optimality gap (*LP OG*) obtained with the first method and reported it for all instances in the table. In addition, we included the percentage of the total CPU time spent in the UM generation portion of the algorithm (*CPU UM*) and the number of UM generated (*Nb.UM*) for the first method, and the CPU time change (*CPU ratio*) and percentage number of short connections change (*Nb.SC chg*) between both methods.

Table 4.6 Results for UM+ECP and improved Benders decomposition approaches[†]

	D95A	B757A	B767R	A320B	D9SA	B767S	D9SB	MD80L	Avg.
UM + ECP									
<i>CPU Total</i>	1.07	0.75	3.43	7.28	45.92	251.13	59.43	763.66	
<i>Nb.SC</i>	50	32	28	51	132	101	181	286	
<i>Cost IP</i>	31,52	37,86	38,70	68,46	81,19	116,33	69,01	97,40	
<i>IP MOG</i>	4.77%	1.48%	0.97%	2.23%	0.87%	1.53%	2.04%	6.26%	2.52%
<i>LP OG</i>	2.38%	0.76%	0.00%	1.58%	0.87%	0.31%	0.88%	3.05%	1.23%
<i>CPU UM</i>	69.2%	76.0%	96.2%	95.6%	97.6%	98.0%	92.8%	89.6%	89.4%
<i>Nb.UM</i>	4	2	10	3	4	10	10	5	6
MP CPO									
<i>CPU Total</i>	1.06	0.99	1.10	2.26	8.23	53.18	17.82	56.00	
<i>Nb.SC</i>	41	18	8	37	94	41	122	165	
<i>Cost IP</i>	30,41	37,60	38,70	67,92	80,48	116,42	68,20	92,18	
<i>IP MOG</i>	1.30%	0.79%	0.97%	1.46%	0.00%	1.60%	0.87%	0.96%	0.99%
<i>CPU ratio</i>	1.01	0.76	3.12	3.22	5.58	4.72	3.34	13.64	4.42
<i>Nb.SC chg</i>	-18.0%	-43.8%	-71.4%	-27.5%	-28.8%	-59.4%	-32.6%	-42.3%	-40.5%

[†] All CPU times are in minutes.

One can notice that UM+ECP solves the smaller instances very rapidly, but that on average 89.4% of the computation time is spent in the UM generation part and that this part tends to grow rapidly with problem size. Therefore, there is a great decrease in total CPU time with the Benders decomposition method for the larger instances. One can also notice that the quality of the IP solutions obtained with UM+ECP for three instances are almost as good or a little better (heuristic branch-and-bound) than with the second method, but that for the other instances, two in particular, the maximum optimality gap is much larger than the one computed with the IP solution of the second method. In fact, the average maximum optimality gap is decreased by more than 50% with the improved Benders decomposition approach. The quality of the IP solutions found with the first method seems to be in line with the quality of the corresponding LP solutions (i.e., instances with large IP maximum optimality gaps also have large LP relaxation optimality gaps). Hence, the difference in quality between the two methods cannot be attributed solely to the heuristic branch-and-bound method used in ECP and in the Benders decomposition approach. Useful short connections for crews must have been missing in the ECP. In addition, the number of short connections used in the solutions obtained with ECP is much larger than with the second method. This is not surprising since included UM columns are the ones that maximize the number of short connections used by aircraft. The solutions obtained with the first method are therefore less robust than the ones obtained with the second method which can also handle restricted connection penalties. The UM+ECP solutions could certainly be improved with more UM generated, but it would come at the cost of a larger total computing time, which is already on average more than four times as much as the time needed to solve the integrated model with the improved Benders decomposition method.

4.4 Conclusion

This paper has introduced a generalized model and a solution methodology for the integrated routing of aircraft and scheduling of crews. The methodology combines Benders decomposition and column generation. The Benders master problem solves the crew scheduling problem whereas the aircraft routing problem is dealt with by the Benders subproblem. The generation of Pareto-optimal Benders cuts proved to be useful in improving the speed of convergence. On test instances containing as many as 700 daily legs, the approach outperforms two recently published methods. When compared to the method proposed by CORDEAU *et al.* (2001), where the Benders master problem solves the aircraft routing problem, the improved approach decreases by a factor of almost 10, on average, the time needed to solve the basic integrated model (IM) without deteriorating the solution quality. It can also solve the larger instances of the generalized integrated model (GIM) in reasonable computing time whereas the former approach cannot. Except for GIM on two of the larger instances, solved in two hours for the first one and five hours for the other, all tests performed took less than an hour. All these times are considered reasonable for tactical planning. Compared with our implementation of the extended crew pairing model proposed by COHN and BARNHART (2003), the approach proposed in this paper finds, in less time, solutions that are both more robust and of better quality. Future developments of our approach will consider the possibility of having some flexibility on the departure times of the flight legs so as to reduce crew costs by slightly modifying the flight schedule.

Acknowledgments

This work was supported by the Québec Government (Fonds pour la Formation de Chercheurs et l'Aide à la Recherche), and the Natural Sciences and Engineering

Research Council of Canada. This support is gratefully acknowledged. Thanks are due to two anonymous referees for their valuable comments.

Bibliography

- BARNHART, C., BOLAND, N.L., CLARKE, L.W., JOHNSON, E.L., NEMHAUSER, G.L. AND SHENOI, R.G. (1998a). Flight string models for aircraft fleeting and routing. *Transportation Science*, 32, 208–220.
- BARNHART, C., COHN, A.M., JOHNSON, E.L., KLABJAN, D., NEMHAUSER, G.L. AND VANCE, P.H. (2003). Airline crew scheduling. In R.W. Hall, ed., *Handbook of Transportation Science (2nd edition)*, pages 517–560. Boston: Kluwer.
- BARNHART, C., LU, F. AND SHENOI, R. (1998c). Integrated airline schedule planning. In *Operations Research in the Airline Industry*, pages 384–403. Boston: Kluwer.
- BENDERS, J.F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4, 238–252.
- CHEBALOV, S. AND KLABJAN, D. (2006). Robust airline crew scheduling: Move-up crews. *Transportation Science*. To be published.
- CLARKE, L.W., HANE, C.A., JOHNSON, E.L. AND NEMHAUSER, G.L. (1996). Maintenance and crew considerations in fleet assignment. *Transportation Science*, 30, 249–260.
- CLARKE, L.W., JOHNSON, E.L., NEMHAUSER, G.L. AND ZHU, Z. (1997). The aircraft rotation problem. *Annals of Operations Research*, 69, 33–46.
- COHN, A.M. AND BARNHART, C. (2003). Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research*, 51, 387–396.

- CORDEAU, J.-F., STOJKOVIĆ, G., SOUMIS, F. AND DESROSIERS, J. (2001). Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation Science*, 35, 375–388.
- DASKIN, M.S. AND PANAYOTOPOULOS, N.D. (1989). A Lagrangian relaxation approach to assigning aircraft to routes in hub and spoke network. *Transportation Science*, 23, 91–99.
- DESAULNIERS, G., DESROSIERS, J., DUMAS, Y., SOLOMON, M.M. AND SOUMIS, F. (1997b). Daily aircraft routing and scheduling. *Management Science*, 43, 841–855.
- DESAULNIERS, G., DESROSIERS, J., IOACHIM, I., SOLOMON, M.M., SOUMIS, F. AND VILLENEUVE, D. (1998b). A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. In T.G. Crainic and G. Laporte, eds, *Fleet Management and Logistics*, pages 57–93. Boston: Kluwer.
- EHRGOTT, M. AND RYAN, D.M. (2002). Constructing robust crew schedules with bicriteria optimization. *Journal of Multi-Criteria Decision Analysis*, 11, 139–150.
- FEO, T.A. AND BARD, J.F. (1989). Flight scheduling and maintenance base planning. *Management Science*, 35, 1415–1432.
- GOPALAN, R. AND TALLURI, K.T. (1998). The aircraft maintenance routing problem. *Operations Research*, 46, 260–271.
- KLABJAN, D., JOHNSON, E.L. AND NEMHAUSER, G.L. (2002). Airline crew scheduling with time windows and plane count constraints. *Transportation Science*, 36, 337–348.
- MAGNANTI, T.L. AND WONG, R.T. (1981). Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29, 464–484.

- NEMHAUSER, G. L. AND WOLSEY, L. A. (1988). *Integer and Combinatorial Optimization*. New York: Wiley.
- RUSHMEIER, R.A. AND KONTOGIORGIS, S.A. (1997). Advances in the optimization of airline fleet assignment. *Transportation Science*, 31, 159–169.
- SUBRAMANIAN, R., JR., R.P. SCHEFF, QUILLINAN, J.D., WIPER, D.S. AND MARSTEN, R.E. (1994). Coldstart: Fleet assignment at Delta air lines. *Interfaces*, 24, 104–120.
- TALLURI, K.T. (1998). The four-day aircraft maintenance routing problem. *Transportation Science*, 32, 43–53.

CHAPITRE 5

AN INTEGRATED AIRCRAFT ROUTING, CREW SCHEDULING AND FLIGHT RETIMING MODEL

Article écrit par Anne Mercier et François Soumis, à paraître dans *Computers & Operations Research*, 2006.

Lorsque l'on considère un niveau supplémentaire d'intégration en permettant de prendre certaines décisions par rapport à l'horaire des vols, la qualité globale des solutions peut être améliorée. Si l'horaire n'est modifié que légèrement, il est raisonnable de croire que la demande pour les vols ne changerait que très peu et que les estimations de coûts demeurent bonnes. Un modèle intégré d'avions et d'équipages tire davantage profit de cette nouvelle flexibilité car les heures de départ sont modifiées en tenant compte de l'effet, non seulement sur les itinéraires d'avions, mais aussi sur les horaires d'équipages et cela serait impossible avec une résolution séquentielle où les modifications faites à un niveau de décision donné pourraient avoir des conséquences fâcheuses au niveau suivant. Évidemment, le nombre de contraintes liantes en est de beaucoup augmenté, ce qui ajoute un degré de difficulté à la résolution du problème.

Le problème intégré de la création d'itinéraires d'avions et d'horaires d'équipages avec possibilités de modifications à l'horaire consiste à déterminer un ensemble de routes d'avions et de rotations d'équipages et à choisir une heure de départ pour chaque segment de vol à l'intérieur d'une fenêtre de temps donnée. Des contraintes liantes assurent que le même horaire soit choisi à la fois pour les routes d'avions et les rotations d'équipages et imposent des temps minimaux de connexion pour les

équipes qui dépendent des connexions utilisées par les appareils ainsi que des heures de départ choisies.

Un modèle de couverture de tâches avec des contraintes liantes de connexions courtes et d'horaire est utilisé pour modéliser le problème intégré. Une contrainte supplémentaire limite le nombre d'avions disponibles. Chaque itinéraire d'avion doit, par ailleurs, respecter une limite sur le nombre d'heures en vol et sur le nombre de jours séparant deux visites à une station d'entretien. Chaque journée de travail dans une rotation d'équipage doit respecter une limite sur sa durée totale, sur le nombre d'heures en vol ainsi que sur le nombre d'atterrissages. De plus, le nombre de journées de travail dans une rotation ne doit pas dépasser une certaine limite. Ces contraintes de réalisabilité des chemins (itinéraires d'avions ou rotations d'équipages) sont modélisées avec l'utilisation de ressources et on peut les traiter directement avec un algorithme de programmation dynamique dans le cadre de la génération de colonnes. Un réseau espace-temps est utilisé pour générer les chemins d'avions et les chemins d'équipages. Dans le réseau d'avions, les arcs représentent soit des arcs de vol ou des arcs de connexion. Dans le réseau d'équipages, chaque arc représente soit un vol travaillé, un vol de repositionnement (*deadhead*), une connexion entre deux vols travaillés, une connexion entre deux vols de repositionnement, ou une connexion entre un vol travaillé et un vol de repositionnement. Les vols de repositionnement permettent aux membres d'équipage de voyager comme passagers lors de certains vols. Les vols de repositionnement permettent souvent une réduction considérable des coûts d'équipages en facilitant le repositionnement d'un équipage vers une ville différente à partir de laquelle l'équipage couvrira de nouveaux segments de vol, ou vers sa base à la fin d'une rotation. Lorsque les modifications à l'horaire sont permises, un ensemble supplémentaire de contraintes doit être inclus pour s'assurer que, pour chaque vol, le même horaire est choisi pour l'équipage qui travaille et celui qui est en vol de repositionnement, s'il y a lieu.

Deux méthodes distinctes peuvent être utilisées pour représenter les fenêtres de temps. On peut, premièrement, inclure des variables qui représentent l'horaire et qui sont contraintes à prendre des valeurs à l'intérieur de l'intervalle choisi ou, deuxièmement, discrétiser ces variables en ne permettant que quelques possibilités de modification à l'horaire. La première méthode a l'avantage de permettre toutes les possibilités donc éventuellement de trouver une meilleure solution, mais elle change la structure du problème en incluant un nouveau type de variable et cela peut en rendre la résolution plus ardue. Cette difficulté additionnelle est souvent injustifié car des petites modifications à l'horaire donnent généralement des solutions de valeur semblable. Nous privilégions donc la deuxième méthode, où seules quelques possibilités sont permises. Dans ce cas, il suffit de mettre autant d'arcs de vol que d'horaires possibles dans les réseaux et une contrainte de couverture, par segment de vol, assure que seulement un de ces arcs soit utilisé dans la solution. La connexion entre deux segments de vol est dite *courte*, pour une combinaison d'heures possibles de départ, si elle est réalisable mais que sa durée est plus petite que le temps de connexion minimal pour les équipages. Dans ce cas, les deux segments de vol ne peuvent être couverts en séquence par le même équipage que si les vols sont aussi couverts par le même appareil. Sinon, la durée de la connexion est insuffisante pour que l'équipage puisse faire le changement d'appareil.

Le modèle intégré avec modifications à l'horaire contient un grand nombre de contraintes liantes de connexion courte qui rendent le problème difficile à résoudre (chaque vol formant une connexion a plusieurs horaires possibles). On peut diminuer le nombre de ces contraintes en les agrégeant de façon à ne conserver qu'une contrainte liante par connexion courte. En fait, par les contraintes de couverture des vols et étant donné que chacune des variables doit être binaire, seul un horaire est choisi pour chaque vol dans les chemins d'avions et seul un horaire est choisi pour chaque vol dans les chemins d'équipages. De plus, par les contraintes liantes d'horaire, chaque vol est associé au même horaire dans les chemins d'avions et d'équipages. Alors, la

même combinaison d'horaires est choisie pour deux vols formant une connexion donnée à la fois dans les chemins d'avions et ceux d'équipages. Ceci implique que, pour une connexion courte donnée, les variables de chemins ne peuvent prendre une valeur différente de zéro que dans une seule des contraintes liantes de connexion courte. On peut facilement montrer que la relaxation linéaire de la formulation agrégée ne l'empêche pas de choisir pour les équipages des fractions de connexions courtes qui sont prises par un avion, mais avec un horaire différent. La formulation agrégée peut donc mener à un plus grand saut d'intégrité. Néanmoins, on peut aussi montrer que non seulement la formulation agrégée est équivalente à la formulation complète en nombres entiers, mais que certaines relaxations linéaires mixtes le sont elles aussi. En effet, par les contraintes liantes d'horaire, le flux sur les arcs d'avions est restreint à être égal au flux sur les arcs d'équipages correspondants. Alors, par exemple, même si les contraintes d'intégrité des variables d'avions sont relaxées, le flux sur les chemins d'avions doit demeurer en nombres entiers (si les contraintes d'intégrité des variables d'équipages sont présentes). Cette propriété sera exploitée par la méthode de résolution.

Un algorithme en trois phases est utilisé pour résoudre le problème intégré avec un horaire fixe. Dans la phase I, toutes les contraintes d'intégrité sont relaxées et la relaxation linéaire est résolue à optimalité par une méthode de décomposition de Benders dans laquelle à la fois le problème-maître d'équipages et le sous-problème d'avions sont résolus par une méthode de génération de colonnes. En conservant toutes les coupes de Benders générées à la phase I, la phase II réintroduit les contraintes d'intégrité pour les variables du problème-maître (chemins d'équipages) et résout le problème résultant en générant des coupes additionnelles. Dans cette phase, le problème-maître en nombres entiers doit être résolu à chaque itération de l'algorithme de décomposition de Benders. Comme les contraintes liantes de connexions courtes dans le sous-problème primal forcent un avion à faire celles qui sont utilisées par les équipages dans le problème-maître, il est possible que le sous-problème primal en

nombres entiers soit non-réalisable à la phase III (pour les chemins d'équipages fixés par le problème-maître) même si le problème d'origine est réalisable. Une étape est donc ajoutée après la troisième phase pour vérifier la réalisabilité du sous-problème en nombres entiers et, si nécessaire, revenir à la phase II pour résoudre de nouveau le problème-maître en nombres entiers avec une contrainte supplémentaire qui empêche le même sous-ensemble de connexions courtes d'être choisi par les équipages.

Une extension directe de cet algorithme ne peut toutefois que très difficilement résoudre le problème intégré avec modifications à l'horaire. Certains raffinements doivent y être ajoutés. Rappelons que le but du sous-problème de Benders est seulement d'évaluer le problème de routage d'avions (ou de vérifier sa réalisabilité) pour un ensemble spécifique de connexions courtes et une heure de départ donnée pour chaque vol. Le sous-problème génère donc des coupes de Benders valides même si une relaxation du problème-maître est utilisée pour lui fournir un ensemble de connexions courtes et un horaire potentiel. Comme il y a un grand nombre de contraintes de coordination d'horaire des vols de repositionnement et que la proportion des vols de repositionnement potentiels qui sont en fait utilisés est petite, on peut générer dynamiquement ces contraintes pour réduire l'effort de calcul requis pour résoudre le problème-maître de Benders. Quand cette approche est utilisée, aucune de ces contraintes n'est incluse au début de la phase I et, après chaque itération de l'algorithme de décomposition de Benders, toutes les contraintes violées sont ajoutées au problème-maître.

Quand le modèle agrégé est utilisé, l'algorithme peut aussi ajouter dynamiquement, durant la phase I, les contraintes liantes de connexion courte complètes qui sont violées pour améliorer la valeur de la relaxation linéaire trouvée à la fin de cette phase. Ajouter des contraintes dans le sous-problème primal de Benders à une itération donnée pendant le processus de résolution n'affecte pas la validité des coupes déjà générées. En fait, restreindre le sous-problème primal implique une relaxation du sous-

problème dual. Par conséquent, les coupes de Benders ajoutées dans les itérations précédant l'ajout des contraintes détaillées ont peut-être été générées à partir de points (ou rayons) intérieurs du polyèdre dual mais néanmoins à partir de solutions duales réalisables du modèle complet. Rappelons que la génération dynamique de ces contraintes n'est pas nécessaire pendant la phase II étant donné que la relaxation linéaire mixte de la formulation agrégée résolue à la phase II est équivalente à la formulation complète à cause de la présence des contraintes liantes d'horaire.

Le modèle intégré avec modifications à l'horaire a été testé avec des données provenant de deux compagnies aériennes de grande taille. La formulation compacte du modèle se résout plus rapidement que la formulation complète, par un ratio d'au moins 5, en moyenne. L'amélioration est principalement attribuable à une diminution du nombre de coupes de Benders générées et non à une diminution du temps de résolution du sous-problème lui-même. Il semble en effet que le sous-problème agrégé génère des coupes de Benders plus fortes que le sous-problème complet. Bien que cela puisse étonner, la formulation compacte détermine une région admissible duale plus petite, ce qui l'aide à générer des coupes plus fortes. Ne mettre les contraintes de coordination des vols de repositionnement dans le problème-maître de Benders que lorsqu'elles sont violées diminue le temps CPU moyen par un autre facteur de 3. Par ailleurs, la génération dynamique des contraintes liantes de connexions courtes désagrégées réduit encore le temps CPU total moyen. Certaines instances ne peuvent d'ailleurs être résolues en temps raisonnable sans ce dernier raffinement. Globalement, la méthode qui combine une formulation compacte du problème et une génération dynamique de coupes diminue le temps CPU total moyen par un facteur de plus de 12 lorsqu'elle est comparée à une implantation directe de la décomposition de Benders.

Les résultats numériques montrent aussi que, pour toutes les instances, le modèle avec modifications à l'horaire produit une solution entière à la fois moins chère pour ce qui est des coûts d'équipages et nécessitant un plus petit nombre d'appareils par rapport à

un modèle avec un horaire fixe. Les coûts variables d'équipages (qui incluent les coûts d'attente et ceux des vols de repositionnement) sont réduits, en moyenne, de 8,3% quand les heures de départ des vols peuvent être avancées ou reculées par aussi peu que cinq minutes. En même temps, ces petites modifications à l'horaire permettent de diminuer de deux, en moyenne, le nombre d'appareils nécessaires à la création d'itinéraires d'avions respectant les contraintes d'entretien.

An Integrated Aircraft Routing, Crew Scheduling and Flight Retiming Model

ANNE MERCIER

École Polytechnique de Montréal

FRANÇOIS SOUMIS

École Polytechnique de Montréal

August 2005

Abstract

In the integrated aircraft routing, crew scheduling and flight retiming problem, a minimum-cost set of aircraft routes and crew pairings must be constructed while choosing a departure time for each flight leg within a given time window. Linking constraints ensure that the same schedule is chosen for both the aircraft routes and the crew pairings, and impose minimum connection times for crews that depend on aircraft connections and departure times. We propose a compact formulation of the problem and a Benders decomposition method with a dynamic constraint generation procedure to solve it. Computational experiments performed on test instances provided by two major airlines show that allowing some flexibility on the departure times within an integrated model yields significant cost savings while ensuring the feasibility of the resulting aircraft routes and crew pairings.

Keywords: aircraft routing; crew scheduling; integrated planning; Benders decomposition; column generation.

Introduction

Airlines usually use a sequential procedure to plan their operations (see, e.g., YU (1998)). By solving a *flight scheduling problem*, they first create a schedule that specifies each flight leg to be flown during a given period and sets departure and arrival times for each of those legs. Then, the *fleet assignment* is performed to assign an aircraft type to each flight leg to maximize anticipated profits while taking into account the number of available aircraft. For each aircraft type, an *aircraft routing problem* is then solved to determine the sequence of flight legs to be flown by each individual aircraft so that each leg is covered exactly once while ensuring appropriate aircraft maintenance. With the aircraft routes on hand, the airline then builds crew rotations or *pairings* by solving a *crew scheduling problem* for each aircraft type. A pairing is a sequence of duty periods separated by overnight rests, and a *duty period* is a sequence of flight legs separated by smaller rest periods, called *sits* (or *connections*). The objective of the crew scheduling problem is to determine a minimum-cost set of pairings so that every flight leg is assigned a qualified crew and every pairing satisfies the set of applicable work rules. For example, each duty period in a pairing must respect limits on total work time, total flight time and total number of landings. In the last step of the planning process, pairings are finally combined to form personalized monthly schedules that are assigned to employees by solving a *crew bidding problem* or a *crew rostering problem*.

While a sequential procedure greatly simplifies the process, CORDEAU *et al.* (2001), KLABJAN *et al.* (2002) and COHN and BARNHART (2003) have shown that integrating the aircraft routing and crew scheduling problems can generate solutions that are significantly better than those obtained by solving the problems sequentially. Because the minimum connection time required between two successive flight legs covered by the same crew depends on whether the same aircraft is used on both legs, aircraft

routing decisions have an impact on the set of feasible pairings. Consequently, a sequential planning procedure is likely to yield suboptimal solutions. A connection that is too short to be made by a crew when the two associated legs are not flown by the same aircraft is said to be *short*. In this paper, we consider an additional level of integration by adding some flight scheduling decisions to the integrated aircraft routing and crew scheduling problem. More precisely, the departure time of each flight leg is allowed to deviate slightly from the planned schedule. Obviously, the same departure time has to be chosen for both the aircraft and the crews, and this complicates the problem. However, an integrated approach can take advantage of the added schedule flexibility to a greater extent since the departure times are chosen by taking into account the benefits to both the aircraft routings and the crew pairings. This would not be possible with a sequential solution process in which modifying the schedule in one step could have unforeseen consequences on the next step. When only small modifications from the original flight schedule are considered, it is reasonable to assume that flight demand does not change significantly (see, e.g., DESAULNIERS *et al.* (1997b), REXING *et al.* (2000) and KLABJAN *et al.* (2002)).

Several modeling and solution approaches have been proposed to separately address the aircraft routing and crew pairing problems. The former problem was studied, among others, by DASKIN and PANAYOTOPOULOS (1989), FEO and BARD (1989), CLARKE *et al.* (1997), GOPALAN and TALLURI (1998) and TALLURI (1998). Numerous contributions regarding the different variants of the crew scheduling problem can also be found in the operations research literature. For an overview, the reader is referred to the recent survey of BARNHART *et al.* (2003). Issues related to the introduction of maintenance and crew considerations in the fleet assignment problem were discussed by CLARKE *et al.* (1996), RUSHMEIER and KONTOGIORGIS (1997) and BARNHART *et al.* (1998c). Finally, other interesting contributions with respect to the integration of the planning process are the approaches presented by

DESAULNIERS *et al.* (1997b) and BARNHART *et al.* (1998a) for the combined fleet assignment and aircraft routing problem.

In recent years, there has been a growing interest in the integration of aircraft routing and crew scheduling problems. CORDEAU *et al.* (2001) have introduced a model that integrates the complete aircraft routing and crew pairing formulations to which is added one linking constraint for each short connection. To handle these linking constraints, a solution approach based on Benders decomposition is used. The solution process iterates between a master problem that solves the aircraft routing problem, and a subproblem that solves the crew pairing problem. Short connections are fixed by the master problem and the subproblem constructs minimum-cost crew pairings using only the fixed set of short connections. Because of their particular structure, both of these problems are solved by column generation. On a set of test instances based on data provided by a Canadian airline, the integrated approach reduced variable crew costs by 9.4% with respect to the sequential planning process commonly used in practice.

The latter model was further enhanced by MERCIER *et al.* (2005) who have introduced a generalized formulation in which solution robustness is improved by penalizing connections that are likely to introduce delays if they are not performed by the same aircraft. The authors also show that reversing the order of the solution sequence, i.e., solving the crew pairing problem in the Benders master problem as opposed to the aircraft routing problem, yields important improvements over the approach of CORDEAU *et al.* (2001). Most costs in the integrated problem are associated with the crew pairings and, by reversing the natural solution sequence, the aircraft routing subproblem is mostly transferring feasibility information to the master problem and very little optimality (or cost) information. This results in a significant decrease in the number of Benders cuts generated. The identification of Pareto-optimal cuts was also shown to be useful in further improving the speed of convergence.

COHN and BARNHART (2003) have also proposed an integrated model, but instead of incorporating the aircraft routing formulation in the model, variables representing complete solutions to the aircraft routing problem are used in an *extended crew pairing model*. This obviously reduces the number of constraints, but may lead to a very large number of additional variables. The authors show that only a subset of the feasible aircraft routing solutions needs to be included in the model, i.e., one column for each *unique and maximal maintenance-feasible short connection set*. These columns can be generated individually and sequentially, in a preprocessing step, by solving a series of aircraft routing problems with additional constraints and a modified objective function. They also propose to solve the extended crew pairing problem by a branch-and-price algorithm in which both crew pairing and aircraft routing solutions are generated dynamically.

KLABJAN *et al.* (2002) have presented a partially integrated approach that solves the crew pairing problem first, but includes additional constraints that count the number of available aircraft on the ground at any time. Under the assumption that maintenance is performed at night when all aircraft are on the ground, these constraints generally ensure the feasibility of the resulting maintenance aircraft routing problem. In addition, the model allows the departure time of each flight leg to be moved within a small time window so as to further reduce crew costs. A rapid depth-first search method generates a subset of pairings that are based on the original schedule and some modified feasibility parameters. For example, the minimum connection time is decreased by the time window width to allow pairings that would be feasible with modified leg departure times. While generating a pairing, if it is impossible to retime the legs in accordance with the true feasibility parameters, the pairing is rejected. To speed up the algorithm, only one feasible retiming of legs per pairing is considered. Next, a crew pairing problem with plane count constraints is solved by considering only the valid generated pairings. Because the time windows

can modify the set of ground arcs, it is difficult to model the plane count constraints exactly and they are thus approximated. On test instances involving up to 450 flight legs, this approach has produced crew solutions with significantly lower costs than the solutions obtained with the traditional sequential method. It is, however, less likely to yield a feasible maintenance aircraft routing problem in an international context where maintenance does not necessarily take place at night.

In the case of the fleet assignment problem, a number of papers have considered the idea of integrating flight scheduling decisions to increase flexibility and ultimately find better solutions (see, e.g., LEVIN (1971), DESAULNIERS *et al.* (1997b) and REXING *et al.* (2000)). However, we are not aware of any model for the fully integrated aircraft routing and crew scheduling problem that also includes flight scheduling decisions. The contributions of this paper are to introduce such a model, to explain how it can be solved efficiently and to evaluate the benefits that result from the increased flexibility related to the departure times. In particular, we show that a straightforward extension of the integrated aircraft routing and crew scheduling model proposed by MERCIER *et al.* (2005) yields an intractable formulation, but that a compact reformulation of the problem coupled with dynamic constraint generation allow the solution of large-scale instances in reasonable computing times.

The remainder of the article is organized as follows. The next section introduces some notation and a mathematical formulation of the problem while Section 5.2 presents the solution methodology. Computational experiments that show the benefits of solving an integrated model including flight retiming are reported in Section 5.3. Conclusions and directions for future work are discussed in the final section.

5.1 Mathematical Formulation

In this paper, we assume that the fleet assignment problem has been solved so that the aircraft type assigned to each flight leg is known. In this context, the integrated aircraft routing, crew scheduling and flight retiming problem decomposes into one problem for each aircraft type. Given a set of flight legs to be flown by the aircraft of a specific type, the problem is then to determine a modified schedule and a minimum-cost set of aircraft routes and crew pairings such that each flight leg is covered by one aircraft and one crew, and side constraints are satisfied.

5.1.1 Model

Our formulation addresses the daily problem which is common in the crew scheduling literature. Consider a set L of daily flight legs to be flown by a single aircraft type. Each flight leg $i \in L$ is defined by origin and destination stations, and departure and arrival times. A finite number of possible changes to the original departure time of a leg is used to model schedule flexibility. Let U_i be the set of possible departure times of leg $i \in L$. For example, if the original departure time of leg i is 12h00, then $U_i = \{11h55, 12h00, 12h05\}$ could be a set of possible departure times for this leg.

Given two flight legs $i, j \in L$, the connection between these legs is said to be short if it is feasible but the difference between the departure time of leg j and the arrival time of leg i is smaller than the minimum sit time for crews. In this case, the legs can be covered, in sequence, by the same crew only if both legs are also covered by the same aircraft. Otherwise, insufficient time is available for the crew to make the connection. Let S be the set of leg pairs for which the connection between them is short for at least one possible combination of departure times. For each $(i, j) \in S$, let

S_{ij} be the set of pairs of departure times $p \in U_i$ and $q \in U_j$ for which the connection between leg i and leg j is short.

The problem is modeled with a path formulation. Each aircraft route must respect a limit on the total number of days separating two visits at a maintenance station. Each duty period in a pairing must respect limits on total work time, total flight time and total number of landings. In addition, the number of duty periods in a pairing must not exceed a certain limit. These path feasibility constraints are modeled through the use of resources and are handled directly by dynamic programming in a column generation framework (see, e.g., DESAULNIERS *et al.* (1998b)). The aircraft and the crew paths are generated with time-space networks. Each node of these networks correspond either to the departure of a flight leg, or to its arrival. Aircraft networks contain two types of arcs: flight arcs and connection arcs. In the crew networks, each arc represents either a flight, a *deadhead* flight, or a feasible connection between two flights, between two deadhead flights or between one of each. Deadheads permit crew members to travel as passengers on certain flights. They are useful to reposition crews to a different city where they are needed to cover a flight leg. They can also be used to ensure that the crew can return to its base at the end of a pairing. For each leg $i \in L$, $|U_i|$ copies of the corresponding flight arc are included in the networks (one for each possible schedule) and the leg covering constraints will ensure that only one of them is used per leg.

Let Ω^F be the set of feasible aircraft paths and let Ω^K denote the set of feasible crew paths. For every path $\omega \in \Omega^F$ or $\omega \in \Omega^K$, define binary constants b_ω^i that take value 1 if leg $i \in L$ belongs to this path and binary constants b_ω^{iu} that take value 1 if leg $i \in L$ is assigned schedule $u \in U_i$ in this path. Let also n_ω^{ijpq} be equal to 1 if leg $i \in L$ with schedule $p \in U_i$ and leg $j \in L$ with schedule $q \in U_j$ are performed in sequence in path ω . Let also c_ω be the cost of sending one unit of flow along path ω . For every aircraft path $\omega \in \Omega^F$, let f_ω be the number of aircraft required to cover path

ω . The value of f_ω may be greater than one since aircraft paths can span more than one day and every leg has to be covered daily. Let also θ_ω be a binary variable that represents the flow on aircraft path ω . For every crew path $\omega \in \Omega^K$, let e_ω be the number of duties in the path. A binary variable χ_ω is defined for every crew path ω , and binary constants d_ω^{iu} take value 1 if leg $i \in L$ with schedule $u \in U_i$ is performed as a deadhead in crew path ω . Finally, constants ζ^F and ζ^D represent the number of available aircraft and a limit on the total number of duties in all crew pairings, respectively. Table 5.1 provides a summary of the notation used in the formulation.

Table 5.1 Summary of notation

L	set of legs
U_i	set of possible departure times for leg i
S	set of pairs of legs for which the connection between them is short
S_{ij}	set of pairs of departure times $p \in U_i$ and $q \in U_j$ for which the connection between leg i and leg j is short
Ω^F	set of feasible aircraft paths
Ω^K	set of feasible crew paths
χ_ω	binary variable representing the flow on crew path ω
θ_ω	binary variable representing the flow on aircraft path ω
b_ω^i	equal to 1 if leg i belongs to path ω
b_ω^{iu}	equal to 1 if leg i with schedule u belongs to path ω
c_ω	cost of sending one unit of flow along path ω
d_ω^{iu}	equal to 1 if deadhead i with schedule u belongs to path ω
e_ω	number of duties in crew path ω
f_ω	number of aircraft required to cover aircraft path ω
n_ω^{ij}	equal to 1 if leg i and leg j are performed in sequence in path ω
n_ω^{ijpq}	equal to 1 if leg i with schedule p and leg j with schedule q are performed in sequence in path ω
ζ^F	number of available aircraft
ζ^D	total number of duties allowed in all crew pairings

The integrated aircraft routing, crew scheduling and flight retiming model, (M1), can be stated as follows:

$$\text{Minimize} \quad \sum_{\omega \in \Omega^K} c_\omega \chi_\omega + \sum_{\omega \in \Omega^F} c_\omega \theta_\omega \quad (5.1)$$

subject to

$$\sum_{\omega \in \Omega^F} b_{\omega}^i \theta_{\omega} = 1 \quad (i \in L) \quad (5.2)$$

$$\sum_{\omega \in \Omega^F} f_{\omega} \theta_{\omega} \leq \zeta^F \quad (5.3)$$

$$\sum_{\omega \in \Omega^K} b_{\omega}^i \chi_{\omega} = 1 \quad (i \in L) \quad (5.4)$$

$$\sum_{\omega \in \Omega^K} e_{\omega} \chi_{\omega} \leq \zeta^D \quad (5.5)$$

$$\sum_{\omega \in \Omega^K} d_{\omega}^{iu} \chi_{\omega} - \sum_{\omega \in \Omega^K} b_{\omega}^{iu} \chi_{\omega} \leq 0 \quad (i \in L, u \in U_i) \quad (5.6)$$

$$\sum_{\omega \in \Omega^K} b_{\omega}^{iu} \chi_{\omega} - \sum_{\omega \in \Omega^F} b_{\omega}^{iu} \theta_{\omega} = 0 \quad (i \in L, u \in U_i) \quad (5.7)$$

$$\sum_{\omega \in \Omega^K} n_{\omega}^{ijpq} \chi_{\omega} - \sum_{\omega \in \Omega^F} n_{\omega}^{ijpq} \theta_{\omega} \leq 0 \quad ((i, j) \in S, (p, q) \in S_{ij}) \quad (5.8)$$

$$\theta_{\omega} \in \{0, 1\} \quad (\omega \in \Omega^F) \quad (5.9)$$

$$\chi_{\omega} \in \{0, 1\} \quad (\omega \in \Omega^K). \quad (5.10)$$

The objective function (5.1) minimizes the sum of all crew scheduling and aircraft routing costs. An approximate crew cost function including piecewise linear waiting costs and deadhead costs is used. Because each flight leg must be covered by exactly one crew, a large portion of total crew costs is fixed. Hence, the only relevant costs considered in these experiments are those that can be reduced by a better planning of crew pairings. Variable expenses are incurred for connections whose duration exceeds a given threshold because crews must then be credited work time even though they are not actually working. Additional accommodation expenses are also incurred when the rest period between successive duties does not take place at the crewbase. For the aircraft routing problem, airlines sometimes take into consideration *through values* that represent the extra revenues obtained by assigning the same aircraft to a pair

of consecutive flight legs (i.e., a *through*) so that passengers flying from the origin of the first leg to the destination of the second leg do not have to change aircraft. Constraints (5.2) and (5.4) require each leg to be covered by exactly one aircraft and one crew, respectively. Constraint (5.3) imposes a limit on the number of available aircraft and constraint (5.5) limits the total number of duties worked. By restricting the number of duties worked, one can increase their duration and make unattractive short duties which would incur charges for the airline. The minimum paid flying time for crews is not included in our approximate crew cost function but sensitivity analysis showed that it is properly replaced by (5.5). Constraints (5.6) ensure that the same schedule is chosen for the working crew and the traveling crew (deadhead), if any. Similarly, constraints (5.7) ensure that, for every leg, the same schedule is chosen for the aircraft and the crew. Finally, constraints (5.8) guarantee that a crew does not change aircraft if, for the chosen schedule, the connection time is too short. These last two groups of constraints ((5.7) and (5.8)) link the aircraft and the crew problems.

5.1.2 A simpler formulation

Model (M1) contains a large number of short connection linking constraints (5.8) which make the problem hard to solve. Indeed, there are potentially $|U_i| \cdot |U_j|$ constraints of this type for each leg pair $(i, j) \in S$ since the connecting flight legs can each have many possible departure times. One can reduce the number of such constraints by aggregating them so as to keep only one linking constraint per short connection. In fact, by constraints (5.2), (5.4), (5.9) and (5.10), only one departure time is chosen for every flight leg in the aircraft paths and only one departure time is chosen for every flight leg in the crew paths. In addition, by constraints (5.7), every leg is associated with the same departure time in both the crew and the aircraft paths.

Therefore, the same combination of departure times is chosen for two given connecting flight legs in both the crew and the aircraft paths. This implies that the path variables can take a non-zero value in only one of the unaggregated linking constraints (5.8) related to a given short connection. These constraints can thus be rewritten as:

$$\sum_{\omega \in \Omega^K} n_{\omega}^{ij} \chi_{\omega} - \sum_{\omega \in \Omega^F} n_{\omega}^{ij} \theta_{\omega} \leq 0 \quad ((i, j) \in S). \quad (5.11)$$

Although the integer aggregated model, (M2), obtained by replacing (5.8) with (5.11) is equivalent to the original formulation, its linear relaxation does not prevent the model to choose for crews fractions of short connections that are taken by aircraft, but with different schedules. The aggregated formulation could therefore lead to a larger integrality gap or introduce a greater number of fractional variables in the LP solution. The detailed short connection linking constraints (5.8) impose that the flow on each short connection arc in the crew networks be smaller than or equal to the flow on the corresponding arc in the aircraft networks. In contrast, the aggregated formulation only requires that the sum of the flows on all arcs corresponding to a given short connection (for all schedule combinations) be smaller or equal in the crew networks. Figure 5.1 shows an example where a solution to the linear relaxation of the aggregated model violates some of the unaggregated constraints (5.8). For ease of exposition, this example only considers two possible departure times for two connecting flight legs where the connection between leg A and leg B is short for all four possible schedule combinations.

One can see in Figure 5.1 that two detailed constraints are violated with this solution. For instance, 0.2 crew makes the short connection between [Leg A - Schedule 1] and [Leg B - Schedule 1] while no aircraft makes the same connection. Nevertheless, one can easily observe that the MIP obtained by relaxing integrality on either crew or aircraft variables in (M2) is equivalent to (M1). If the aircraft variables, for example,

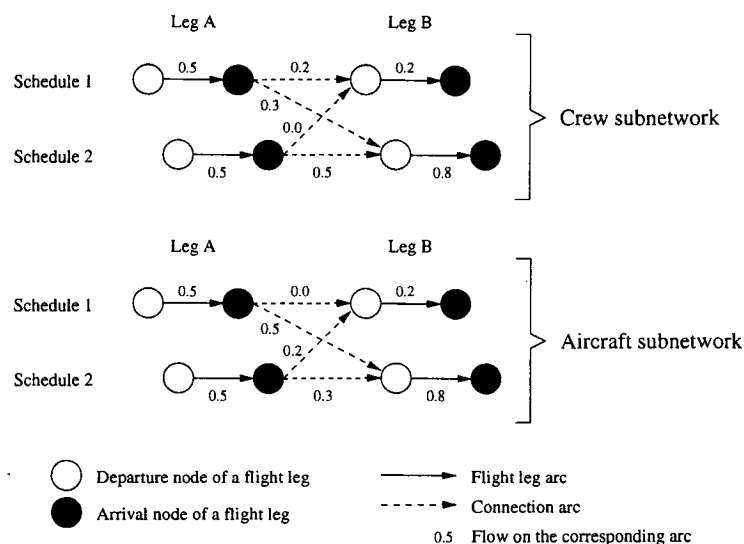


Figure 5.1 Example of some violated unaggregated constraints

are not restricted to take integer values, the flow on the leg arcs would still be integer because of constraints (5.7), which restrict the flow on the aircraft arcs to be equal to the flow on the crew arcs. Consequently, as in the integer formulation of (M2), only one possible schedule combination can be taken between two flight legs.

5.2 Solution Methodology

5.2.1 Benders decomposition

The model includes both crew pairing and aircraft routing path variables. Benders decomposition (see BENDERS (1962)) can be used to reformulate the problem to separate the two types of variables at the expense of an increase in the number of constraints.

Let χ be the set of solutions (paths) satisfying the crew constraints (5.4), (5.5), (5.6) and (5.10). For given integer values $\bar{\chi}_\omega$ ($\omega \in \Omega^K$) $\in \chi$, the MIP relaxation of the aggregated model (obtained by relaxing integrity on variables θ_ω in (M2)) reduces to the following *primal subproblem* involving only aircraft variables:

$$\text{Minimize } \sum_{\omega \in \Omega^F} c_\omega \theta_\omega \quad (5.12)$$

subject to

$$\sum_{\omega \in \Omega^F} b_\omega^i \theta_\omega = 1 \quad (i \in L) \quad (5.13)$$

$$\sum_{\omega \in \Omega^F} f_\omega \theta_\omega \leq \zeta^F \quad (5.14)$$

$$\sum_{\omega \in \Omega^F} b_\omega^{iu} \theta_\omega \geq \sum_{\omega \in \Omega^K} b_\omega^{iu} \bar{\chi}_\omega \quad (i \in L, u \in U_i) \quad (5.15)$$

$$\sum_{\omega \in \Omega^F} n_\omega^{ij} \theta_\omega \geq \sum_{\omega \in \Omega^K} n_\omega^{ij} \bar{\chi}_\omega \quad ((i, j) \in S) \quad (5.16)$$

$$\theta_\omega \geq 0 \quad (\omega \in \Omega^F). \quad (5.17)$$

One can notice that the set of equalities (5.7) has been replaced with inequalities (5.15) in the subproblem. This form is equivalent but easier to solve because it reduces the feasible set of the dual.

Let $\alpha = (\alpha_i | i \in L)$, $\beta \leq 0$, $\delta = (\delta_{iu} \geq 0 | i \in L, u \in U_i)$ and $\mu = (\mu_{ij} \geq 0 | (i, j) \in S)$ be the dual variables associated with constraints (5.13)-(5.16), respectively. The dual of (5.12)-(5.17) is the following *dual subproblem*:

$$\text{Maximize } \sum_{i \in L} \alpha_i + \zeta^F \beta + \sum_{i \in L} \sum_{u \in U_i} \sum_{\omega \in \Omega^K} b_\omega^{iu} \bar{\chi}_\omega \delta_{iu} + \sum_{(i, j) \in S} \sum_{\omega \in \Omega^K} n_\omega^{ij} \bar{\chi}_\omega \mu_{ij} \quad (5.18)$$

subject to

$$\sum_{i \in L} b_{\omega}^i \alpha_i + f_{\omega} \beta + \sum_{i \in L} \sum_{u \in U_i} b_{\omega}^{iu} \delta_{iu} + \sum_{(i,j) \in S} n_{\omega}^{ij} \mu_{ij} \leq c_{\omega} \quad (\omega \in \Omega^F) \quad (5.19)$$

$$\beta \leq 0 \quad (5.20)$$

$$\delta_{iu} \geq 0 \quad (i \in L, u \in U_i) \quad (5.21)$$

$$\mu_{ij} \geq 0 \quad ((i, j) \in S). \quad (5.22)$$

Assuming that $c_{\omega} \geq 0$ for all $\omega \in \Omega^F$, the dual subproblem is always feasible since the null vector $\mathbf{0}$ satisfies constraints (5.19)-(5.22). Furthermore, if it is also bounded, it makes the primal subproblem feasible and bounded as well. Let Δ denote the polyhedron defined by constraints (5.19)-(5.22), and let P_{Δ} and R_{Δ} be the sets of extreme points and extreme rays of Δ , respectively. One can see that Δ does not depend on the crew problem since the crew elements $\bar{\chi}_{\omega}$ are present only in the objective function (5.18). P_{Δ} and R_{Δ} could then be enumerated a priori.

Introducing the additional free variable z_0 , the MIP relaxation of (M2) can thus be reformulated as the following *Benders master problem*:

$$\text{Minimize} \quad \sum_{\omega \in \Omega^K} c_{\omega} \chi_{\omega} + z_0 \quad (5.23)$$

subject to

$$z_0 - \sum_{i \in L} \sum_{u \in U_i} \sum_{\omega \in \Omega^K} b_{\omega}^{iu} \delta_{iu} \chi_{\omega} - \sum_{(i,j) \in S} \sum_{\omega \in \Omega^K} n_{\omega}^{ij} \mu_{ij} \chi_{\omega} \geq \sum_{i \in L} \alpha_i + \zeta^F \beta \quad ((\alpha, \delta, \mu, \beta) \in P_{\Delta}) \quad (5.24)$$

$$- \sum_{i \in L} \sum_{u \in U_i} \sum_{\omega \in \Omega^K} b_{\omega}^{iu} \delta_{iu} \chi_{\omega} - \sum_{(i,j) \in S} \sum_{\omega \in \Omega^K} n_{\omega}^{ij} \mu_{ij} \chi_{\omega} \geq \sum_{i \in L} \alpha_i + \zeta^F \beta \quad ((\alpha, \delta, \mu, \beta) \in R_{\Delta}) \quad (5.25)$$

$$\sum_{\omega \in \Omega^K} b_{\omega}^i \chi_{\omega} = 1 \quad (i \in L) \quad (5.26)$$

$$\sum_{\omega \in \Omega^K} e_{\omega} \chi_{\omega} \leq \zeta^D \quad (5.27)$$

$$\sum_{\omega \in \Omega^K} d_{\omega}^{iu} \chi_{\omega} - \sum_{\omega \in \Omega^K} b_{\omega}^{iu} \chi_{\omega} \leq 0 \quad (i \in L, u \in U_i) \quad (5.28)$$

$$\chi_{\omega} \in \{0, 1\} \quad (\omega \in \Omega^K). \quad (5.29)$$

Feasibility constraints (5.25) ensure that the values given to the crew variables χ_{ω} ($\omega \in \Omega^K$) lead to a bounded dual subproblem. When bounded, the purpose of the dual subproblem is to evaluate the aircraft routing problem for a specific set of short connections and a specific schedule for every leg. The value of z_0 is thus restricted to be larger than or equal to the optimal value of the dual subproblem by optimality constraints (5.24). Since the Benders cuts are generated from the polyhedron of the dual subproblem, integrality on the primal subproblem variables has to be relaxed. However, an algorithm in three phases ensuring that an integer solution to the problem is obtained is described in the next section.

In general, model (5.23)-(5.29) contains more constraints than the MIP relaxation of (M2) but most optimality and feasibility cuts are inactive in an optimal solution.

Hence, these constraints need not be enumerated exhaustively but can instead be generated dynamically by iterating between a relaxed master problem and the subproblem. The relaxed master problem contains constraints (5.26)-(5.29) as well as subsets of Benders cuts (5.24) and (5.25). The optimal solution of the relaxed Benders master problem is used to set up constraints (5.15) and (5.16) in the primal subproblem at every iteration. If the primal subproblem is feasible, the value of the dual variables associated with constraints (5.13)-(5.16) determine an extreme point of P_{Δ} . Otherwise, an extreme ray of R_{Δ} violating one of the constraints (5.25) is identified. When through values are not considered, as it is often the case in the literature (see, e.g., KLABJAN (2005)), the aircraft routing problem reduces to a feasibility problem and optimality cuts become irrelevant. Hence, exactly one feasibility cut is added to the relaxed Benders master problem at each iteration and the process continues until its optimal solution yields a feasible primal subproblem.

5.2.1.1 Three-phase algorithm

MERCIER *et al.* (2005) have described a heuristic solution method in which both the crew pairing master problem and the aircraft routing subproblem are solved by column generation. The method works in three phases. In Phase I, all integrality requirements are relaxed and the relaxation is solved to optimality by Benders decomposition and column generation. Retaining all generated cuts, Phase II reintroduces integrality constraints on the master problem crew variables and solves the resulting mixed-integer problem by generating additional cuts. In this phase, the integer master problem must be solved at each iteration of the Benders decomposition algorithm. In Phase III, integrality constraints are finally added on the subproblem aircraft variables and the integer subproblem is solved once with the values of the master problem variables being held fixed. Because linking constraints in the primal subproblem force

the aircraft to use the short connections selected for crews in the master problem, the integer primal subproblem may be infeasible in Phase III for the given solution of the master problem, even if the original problem is feasible. A step is therefore added after the third phase to verify the feasibility of the integer subproblem and, if needed, go back to the second phase to solve the integer master problem with an additional constraint forbidding the same subset of short connections to be chosen. To obtain integer solutions in the master problem and in the subproblem, a heuristic branching strategy is used. Branching is performed on the path variables and decisions can be made simultaneously on more than one variable to accelerate the search.

5.2.1.2 Dynamic constraint generation

Due to the increased number of constraints and variables in the model to include possible schedule changes, the algorithm developed by MERCIER *et al.* (2005) does not succeed in solving the model efficiently. In fact, the computational experiments presented in Section 5.3 show that the straightforward extension of the method does not converge within 36 hours of computing time for the larger instances, even when the aggregated model is used. The integrated model including flight retiming is thus solved with a modified version of the three-phase approach. The new approach includes a dynamic constraint generation procedure.

The large number of deadhead coordination constraints (5.6) in the model contributes to the inefficiency of the solution method, but without deadhead flights, the majority of the instances would either become infeasible or the crew costs would increase significantly. However, since the proportion of potential deadhead flights actually used in the solutions is small, constraints (5.6) can be generated dynamically to reduce the computational effort required to solve the Benders master problem. When this approach is used, all constraints (5.6) are relaxed in the master problem at the

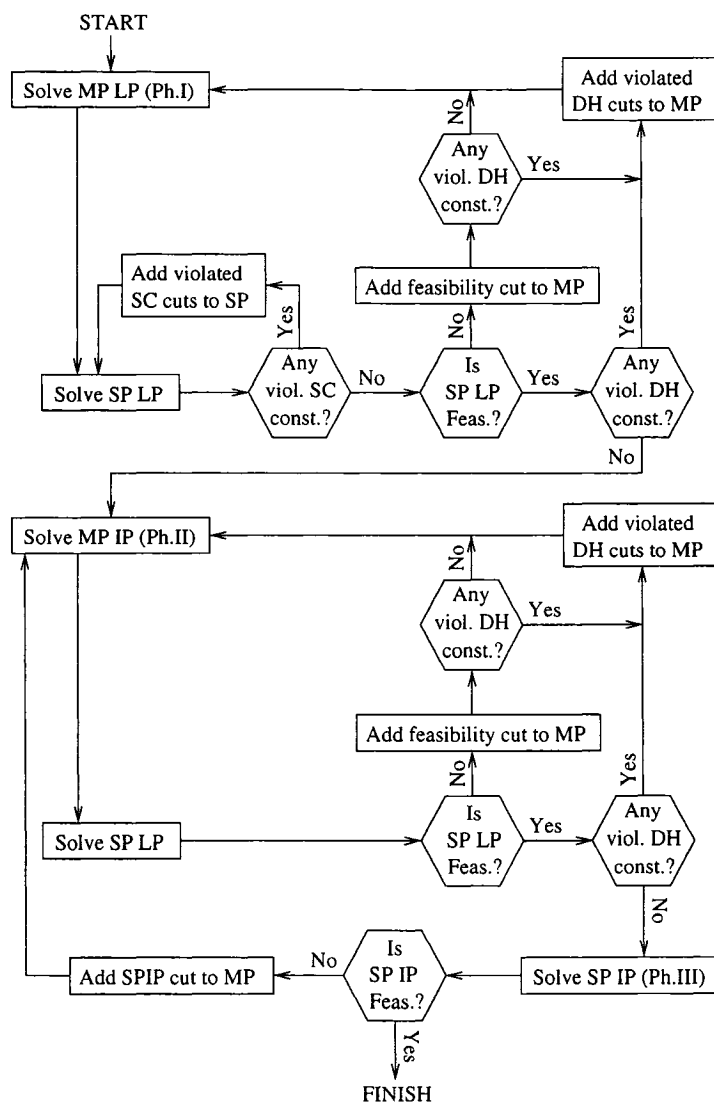
beginning of phase I and, after each iteration of the Benders decomposition algorithm, all violated constraints are added to the master problem. Recall that the purpose of the Benders subproblem is only to evaluate the aircraft routing problem (or verify its feasibility) for a specific set of short connections and a specific schedule for every leg. The subproblem thus generates valid Benders cuts even if a relaxation of the master problem is used to supply crew pairings and potential schedules.

When the aggregated model is used, the algorithm can also add, throughout phase I, the violated detailed short connection linking constraints (5.8) to the model in an effort to improve the LP solution value. Adding constraints in the primal subproblem during the solution process does not affect the validity of the previously generated Benders cuts. Indeed, restricting the primal subproblem relaxes the dual subproblem. As a result, Benders cuts added in previous iterations were generated from possibly interior points (rays) of the dual, but nonetheless feasible dual points (rays) of the complete model. In other words, the missing dual values (associated with the detailed linking constraints) were assumed to be 0 in the previous iterations. Recall that this step is not needed in phase II since the MIP relaxation of (M2) is not a relaxation of the integer formulation of (M1) because of constraints (5.7).

Figure 5.2 gives the complete algorithmic flowchart of the proposed solution method.

5.3 Computational Experiments

In this section, we present computational experiments that were carried out on a set of seven instances based on data provided by two major airlines. Three possible departure times for every leg, five minutes apart, were used to test the flight retiming aspect and make comparisons with the model in which the schedule is fixed. These



MP LP	Linear relaxation of the master problem
SP LP	Linear relaxation of the subproblem
MP IP	Integral master problem
SP IP	Integral subproblem
viol. SC const.	Violated detailed short connection linking constraints
viol. DH const.	Violated deadhead schedule coordination constraints
SPIP cut	Forbidden short connection subset cut

Figure 5.2 Algorithmic flowchart

modifications from the original flight schedule are considered small enough not to change flight demand.

5.3.1 Description of data sets

The test instances come from daily fleet assignment solutions provided by the airlines. Some characteristics of the different instances are given in Table 5.2. This table indicates, for each instance, the number of daily legs and the total number of possible short connections ($|S|$). In column $|S'|$, we indicate the number of connections that are short for at least one combination of departure times when small retimings are considered (± 5 min.). The percentage increase in the number of possible short connections with the retimings is given in the last column of the table.

Table 5.2 Characteristics of test instances

Instance	Legs	$ S $	$ S' $	Increase
D95A	198	87	127	46.0%
B757A	184	114	151	32.5%
B767R	152	83	168	102.4%
A320D	258	182	272	49.5%
D9SA	523	502	725	44.4%
D9SB	508	659	961	45.8%
B767S	510	370	697	88.4%
Avg.				58.4%

One can see from Table 5.2 that the flight retiming model contains, on average, 58.4% more possible short connections than the model with fixed departure times. Small schedule modifications can thus have a strong impact on the number of pairings satisfying all the work rules and on the number of maintenance-feasible aircraft routes. Allowing new paths that would otherwise be infeasible can significantly reduce crew costs while reducing the number of necessary aircraft.

5.3.2 Summary of computational experiments

To evaluate the efficiency of the three-phase algorithm on the integrated model with flight retiming, we first tried to solve the model where all the detailed short connection linking constraints and all the deadhead schedule coordination constraints are included from the start (model (M1)). For more than half of the instances, this straightforward extension could not find a solution even after 36 hours of computing time. To improve the efficiency of the algorithm, we tried to solve the compact aggregated model, first with all the deadhead coordination constraints (model (M2)), and then with the latter constraints generated dynamically (model (M2a)). Finally, we tried solving the compact model with the deadhead coordination constraints generated dynamically again but also with the violated detailed short connection linking constraints generated as needed throughout phase I (model (M2b)). Our algorithms were coded in C++ and use GENCOL¹ for column generation. All experiments were performed on a Pentium 4, 2 GHz computer, using a single processor.

Table 5.3 presents a comparison of the CPU time and computational effort needed to perform all three phases with the different approaches. We indicate, for the four approaches and all seven instances, the time spent in each of the three phases as well as the number of cuts generated in the first two phases and the number of forbidden short connection subset cuts added after phase II to get a feasible integer subproblem (*SPIP Cuts*). We also indicate, when appropriate, the total number of deadhead schedule coordination constraints added (*DH Cuts*) and the total number of detailed short connection linking constraints added (*SCLC Cuts*). *Cost IP* indicates the cost of the integer integrated solution at the end of Phase III. Since the branch-and-bound strategy is heuristic, the optimality of the solution is not guaranteed. The master problem also includes an integrality gap. The reported gap is therefore the maximum

¹GENCOL is an optimization software that was developed at GERAD in Montreal.

optimality gap and is computed as the percentage difference between *Cost IP* and *Cost LP*. Finally, the CPU time efficiency of the methods are compared. For example, *CPU ratio vs M1* reported for (M2a) is the total CPU time of (M1) divided by the total CPU time of (M2a).

When considering Phase I CPU times, one observes that the approach using the compact aggregated formulation (M2) always produced an LP solution faster than the one using the unaggregated model (M1). In addition, the LP value does not deteriorate with the aggregated model. Although the number of linking constraints is greatly reduced in the aggregated Benders subproblem, the improvements are mainly attributed to a decrease in the number of Benders cuts generated and not to the improved solution time of the subproblem itself. It thus seems that the aggregated subproblem generates stronger Benders cuts than the unaggregated one. While this may be surprising, it can be explained by the fact that the aggregated model has a smaller dual feasible region which helps to generate stronger cuts. The comparison of the time spent in finding an integer solution can be misleading because of the heuristic branch-and-bound method. For all instances, (M2) nevertheless solved the integer problem faster than (M1), by a factor of at least 5.10, on average, and an additional instance could be solved within the time limit. This CPU decrease is in fact underestimating the real ratio of decrease since (M1) was sometimes stopped before getting a solution.

Putting the deadhead schedule coordination constraints in the crew master problem only when they are violated (model (M2a)) further improves the performance of the three-phase algorithm. This refinement can decrease the average total CPU time by another factor of at least 3.30 when compared to the basic aggregated model (M2) and two additional instances could be solved. In addition to reducing the solution time of the crew pairing problem, the numerical results show that this relaxation of

Table 5.3 Complete results: integrated problem with flight retiming[†]

	D95A	B757C	B767R	A320D	D9SA	D9SB	B767S	Avg.
<i>Cost LP</i>	20,626	47,821	37,492	80,272	66,840	80,229	172,728	
<i>M1</i>								
<i>CPU Ph.I</i>	>36h	41.51	87.40	153.49	>36h	>36h	>36h	
<i>CPU Ph.II</i>		19.44	7.26	8.80				
<i>CPU Ph.III</i>		34.34	20.72	1.85				
<i>CPU Total</i>	>36h	95.29	115.38	164.14	>36h	>36h	>36h	
<i>Cuts Ph.I</i>		68	69	78				
<i>Cuts Ph.II</i>		6	0	0				
<i>SPIP Cuts</i>		2	2	0				
<i>Cost IP</i>		48,973	37,543	80,848				
<i>Gap</i>		2.35%	0.14%	1.93%				1.47%
<i>M2</i>								
<i>CPU Ph.I</i>	158.80	27.81	9.25	46.62	>36h	>36h	>36h	
<i>CPU Ph.II</i>	17.08	27.05	2.68	29.42				
<i>CPU Ph.III</i>	0.17	3.65	14.58	0.32				
<i>CPU Total</i>	176.05	58.51	26.51	76.36	>36h	>36h	>36h	
<i>Cuts Ph.I</i>	95	34	3	31				
<i>Cuts Ph.II</i>	2	12	0	4				
<i>SPIP Cuts</i>	0	1	0	0				
<i>Cost IP</i>	20,748	48,520	37,543	81,828				
<i>Gap</i>	0.59%	1.44%	0.14%	1.90%				1.02%
<i>CPU ratio vs M1</i>	>12.27	1.63	4.35	2.15				>5.10
<i>M2a</i>								
<i>CPU Ph.I</i>	32.30	8.75	10.61	14.64	316.63	272.10	58.01	
<i>CPU Ph.II</i>	9.98	3.79	1.44	7.51	12.89	181.67	>36h	
<i>CPU Ph.III</i>	0.14	2.97	14.63	1.17	32.08	868.93		
<i>CPU Total</i>	42.42	15.51	26.68	23.32	361.60	1322.70	>36h	
<i>Cuts Ph.I</i>	17	8	6	11	16	9	4	
<i>Cuts Ph.II</i>	2	1	0	1	0	3	>3	
<i>SPIP Cuts</i>	0	0	0	0	0	1		
<i>DH cuts</i>	0	7	3	10	0	2		
<i>Cost IP</i>	20,688	48,348	37,543	80,407	67,405	81,123		
<i>Gap</i>	0.30%	1.09%	0.14%	0.17%	0.84%	1.10%		0.61%
<i>CPU ratio vs M2</i>	4.15	3.77	0.99	3.27	>5.97	>1.63		>3.30
<i>M2b</i>								
<i>CPU Ph.I</i>	29.83	7.61	12.17	13.19	436.09	269.90	37.26	
<i>CPU Ph.II</i>	10.32	0.99	1.27	1.78	21.74	364.99	671.12	
<i>CPU Ph.III</i>	0.44	2.28	5.90	4.18	43.00	711.63	211.34	
<i>CPU Total</i>	40.59	10.88	19.34	19.15	500.83	1346.52	919.72	
<i>Cuts Ph.I</i>	18	12	6	6	20	9	3	
<i>Cuts Ph.II</i>	2	0	0	0	0	5	0	
<i>SPIP Cuts</i>	0	0	0	0	0	1	0	
<i>SCLC cuts</i>	9	7	8	11	2	24	28	
<i>Cost IP</i>	20,748	48,112	37,555	80,388	66,845	81,157	176,975	
<i>Gap</i>	0.59%	0.60%	0.17%	0.14%	0.01%	1.14%	2.40%	0.72%
<i>CPU ratio vs M2a</i>	1.05	1.43	1.38	1.22	0.72	0.98	>2.35	>1.34
<i>CPU ratio vs M1</i>	>53.22	8.76	5.97	8.57	>4.31	>1.60	>2.35	>12.11

[†] All CPU times are in minutes.

the Benders master problem also has the effect of reducing the number of iterations of the Benders decomposition algorithm.

Furthermore, the dynamic generation of the violated detailed short connection linking constraints in the course of phase I (model (M2b)) can also decrease the average total CPU time. This refinement is essential to be able to solve all the larger instances. Recall that the LP values of (M2) are equal to the LP values of (M1) on these instances. Therefore, adding the detailed short connection linking constraints cannot improve the LP values, but it can still speed up phase II since restoring integrality on the crew path variables in this phase implies the satisfaction of the detailed linking constraints in the subproblem, on account of constraints (5.7). Indeed, one can observe that the number of generated Benders cuts in phase II of (M2b) is always lower than or equal to the number generated with (M2a), except for instance D9SB. For instance D9SB, some variability may have come from the generation of an SPIP cut to get a feasible integer subproblem. Yet, it is worth noting that the number of Benders cut generated in the first run of phase II (before the SPIP cut) was again lower with (M2b). When this approach is compared to the straightforward extension (model (M1)), the total CPU time is decreased, on average, by a factor of more than 12. One can finally notice that the performance improvements do not come at the price of lower solution quality. When comparing (M2b) to any other method, if the average maximum optimality gap is computed only on the instances solvable with the two methods, the gap is always lower when using (M2b).

It is worth noting that we also tried to generate Pareto-optimal Benders cuts (see, e.g., MAGNANTI and WONG (1981)) but the time needed to solve the auxiliary problem used to generate non-dominated cuts was always greater than the time saved by having a reduced number of iterations of the Benders decomposition procedure.

To evaluate the benefits of having some flexibility in the schedule, the results of (M2b) (the most efficient approach) were compared with the results obtained when solving the integrated model with a fixed schedule (model (M0)). Table 5.4 compares the computational effort needed as well as the quality of the solutions obtained (in terms of costs). We indicate for both approaches the time spent in each of the three phases as well as the number of cuts generated in the first two phases and the number of forbidden short connection subset cuts added after phase II to get a feasible integer subproblem (*SPIP Cuts*). *Cost LP* indicates the cost of the solution found at the end of Phase I and *Cost IP* indicates the cost of the integer solution at the end of Phase III. The maximum optimality gap reported is computed with respect to the corresponding LP value. The quality of the solutions are compared by means of the LP cost percentage decrease (*Cost LP % dec.*), the IP cost percentage decrease (*Cost IP % dec.*) and the reduction in the number of aircraft needed (*Aircraft nb. dec.*). Finally, the CPU times used by the methods are compared (*CPU ratio vs M0*).

These results show that for all instances, the model with flight retiming produced an integer solution of lower cost and with a smaller number of aircraft than the solution produced by the model with fixed departure times. Crew costs, which include waiting costs and deadhead costs, are decreased, on average, by 8.30% when the departure time of the flights can be moved forward or backward by just five minutes. Since these crew costs account, on average, for about 20% of total crew costs (which include a large fixed cost for the actual flight time), the integrated model with flight retiming can reduce the total crew costs by an average of 1.60% in our instances. At the same time, with these small schedule modifications, the number of aircraft needed to respect the maintenance requirements can be reduced, on average, by almost 2. One can notice that the instances for which the crew costs are decreased by a smaller percentage are also the instances for which the number of aircraft was reduced more. Globally, these small schedule modifications can thus significantly reduce airline costs.

Table 5.4 Computational effort and solution quality: with and without flight retiming[†]

	D95A	B757C	B767R	A320D	D9SA	D9SB	B767S	Avg.
<i>M0</i>								
<i>CPU Ph.I</i>	0.52	0.47	0.23	1.03	5.38	7.72	2.33	
<i>CPU Ph.II</i>	0.23	3.29	0.12	0.73	0.14	16.59	2.92	
<i>CPU Ph.III</i>	0.02	0.35	0.23	0.01	1.83	25.07	3.00	
<i>CPU Total</i>	0.77	4.11	0.58	1.77	7.35	49.38	8.25	
<i>Cuts Ph.I</i>	11	14	0	20	9	11	4	
<i>Cuts Ph.II</i>	0	2	0	1	0	2	0	
<i>SPIP Cuts</i>	0	1	0	0	0	3	0	
<i>Cost LP</i>	25,036	49,355	39,690	83,057	80,480	85,481	177,676	
<i>Cost IP</i>	25,403	50,420	40,638	83,367	80,480	86,317	178,980	
<i>Gap</i>	1.44%	2.11%	2.33%	0.37%	0.00%	0.97%	0.73%	1.14%
<i>M2b</i>								
<i>CPU Ph.I</i>	29.83	7.61	12.17	13.19	436.09	269.90	37.26	
<i>CPU Ph.II</i>	10.32	0.99	1.27	1.78	21.74	364.99	671.12	
<i>CPU Ph.III</i>	0.44	2.28	5.90	4.18	43.00	711.63	211.34	
<i>CPU Total</i>	40.59	10.88	19.34	19.15	500.83	1346.52	919.72	
<i>Cuts Ph.I</i>	18	12	6	6	20	9	3	
<i>Cuts Ph.II</i>	2	0	0	0	0	5	0	
<i>SPIP Cuts</i>	0	0	0	0	0	1	0	
<i>Cost LP</i>	20,626	47,821	37,492	80,272	66,840	80,229	172,728	
<i>Cost IP</i>	20,748	48,112	37,555	80,388	66,845	81,157	176,975	
<i>Gap</i>	0.59%	0.60%	0.17%	0.14%	0.01%	1.14%	2.40%	0.72%
<i>Cost LP % dec.</i>	17.61%	3.11%	5.54%	3.35%	16.95%	6.14%	2.78%	7.93%
<i>Cost IP % dec.</i>	18.32%	4.58%	7.59%	3.57%	16.94%	5.98%	1.12%	8.30%
<i>Aircraft nb. dec.</i>	1	2	1	4	1	2	2	1.83
<i>CPU ratio vs M0</i>	52.71	2.65	33.34	10.82	68.14	27.27	111.48	43.77

[†] All CPU times are in minutes.

Of course, the CPU ratios show that the integrated problem with flight retiming is much harder to solve, but the times can still be considered reasonable for tactical planning.

Finally, it is worth mentioning that the benefits displayed in the above table are solely attributable to the introduction of flight retiming in the integrated aircraft routing and crew scheduling problem. One could be interested in comparing the results from the integrated aircraft routing, crew scheduling and flight retiming problem with a sequential procedure also incorporating flight retiming. On the one hand, it can be observed that the number of Benders feasibility cuts is always positive in phase I of (M2b). A sequential procedure with flight retiming that solves the

crew pairing problem first would thus always lead to an infeasible aircraft routing problem. On the other hand, if the aircraft routing problem was solved first, the flight scheduling decisions would be made on a feasibility problem and not with the objective of minimizing crew costs. As a result, the cost decrease would not be as important. Finally, all benefits attributable to the integration of aircraft routing and crew scheduling (see, e.g., CORDEAU *et al.* (2001), KLABJAN *et al.* (2002) and COHN and BARNHART (2003)) would be left out.

5.4 Conclusion

This paper has introduced a model and a solution methodology for the integrated aircraft routing, crew scheduling and flight retiming problem. The methodology combines Benders decomposition, column generation and a dynamic constraint generation procedure. On test instances containing up to 500 daily legs, the approach yields solutions that significantly decrease crew costs while also reducing the number of aircraft and still ensuring appropriate aircraft maintenance. This would not be possible with a sequential solution process. When compared to a straightforward extension of the solution methodology previously developed by MERCIER *et al.* (2005), by aggregating some of the short connection linking constraints in the Benders subproblem and by generating some other constraints dynamically, the new approach decreases by a factor of more than 12, on average, the time needed to solve the integrated model with flight retiming without deteriorating the solution quality.

Acknowledgments

This work was supported by the Québec Government (Fonds pour la Formation de Chercheurs et l'Aide à la Recherche), and the Natural Sciences and Engineering Research Council of Canada. This support is gratefully acknowledged. Thanks are also due to AD OPT Technologies for providing the data used in the computational experiments.

Bibliography

- BARNHART, C., BOLAND, N.L., CLARKE, L.W., JOHNSON, E.L., NEMHAUSER, G.L. AND SHENOI, R.G. (1998a). Flight string models for aircraft fleeting and routing. *Transportation Science*, 32, 208–220.
- BARNHART, C., COHN, A.M., JOHNSON, E.L., KLABJAN, D., NEMHAUSER, G.L. AND VANCE, P.H. (2003). Airline crew scheduling. In R.W. Hall, ed., *Handbook of Transportation Science (2nd edition)*, pages 517–560. Boston: Kluwer.
- BARNHART, C., LU, F. AND SHENOI, R. (1998c). Integrated airline schedule planning. In *Operations Research in the Airline Industry*, pages 384–403. Boston: Kluwer.
- BENDERS, J.F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4, 238–252.
- CLARKE, L.W., HANE, C.A., JOHNSON, E.L. AND NEMHAUSER, G.L. (1996). Maintenance and crew considerations in fleet assignment. *Transportation Science*, 30, 249–260.
- CLARKE, L.W., JOHNSON, E.L., NEMHAUSER, G.L. AND ZHU, Z. (1997). The aircraft rotation problem. *Annals of Operations Research*, 69, 33–46.

- COHN, A.M. AND BARNHART, C. (2003). Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research*, 51, 387–396.
- CORDEAU, J.-F., STOJKOVIĆ, G., SOUMIS, F. AND DESROSIERS, J. (2001). Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation Science*, 35, 375–388.
- DASKIN, M.S. AND PANAYOTOPOULOS, N.D. (1989). A Lagrangian relaxation approach to assigning aircraft to routes in hub and spoke network. *Transportation Science*, 23, 91–99.
- DESAULNIERS, G., DESROSIERS, J., DUMAS, Y., SOLOMON, M.M. AND SOUMIS, F. (1997b). Daily aircraft routing and scheduling. *Management Science*, 43, 841–855.
- DESAULNIERS, G., DESROSIERS, J., IOACHIM, I., SOLOMON, M.M., SOUMIS, F. AND VILLENEUVE, D. (1998b). A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. In T.G. Crainic and G. Laporte, eds, *Fleet Management and Logistics*, pages 57–93. Boston: Kluwer.
- FEO, T.A. AND BARD, J.F. (1989). Flight scheduling and maintenance base planning. *Management Science*, 35, 1415–1432.
- GOPALAN, R. AND TALLURI, K.T. (1998). The aircraft maintenance routing problem. *Operations Research*, 46, 260–271.
- KLABJAN, D. (2005). Large-scale models in the airline industry. In M.M. Solomon G. Desaulniers, J. Desrosiers, ed., *Column Generation*, pages 163–195. New York: Springer.
- KLABJAN, D., JOHNSON, E.L. AND NEMHAUSER, G.L. (2002). Airline crew scheduling with time windows and plane count constraints. *Transportation Science*, 36, 337–348.

- LEVIN, A. (1971). Scheduling and fleet routing models for transportation systems. *Transportation Science*, 5, 232–255.
- MAGNANTI, T.L. AND WONG, R.T. (1981). Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29, 464–484.
- MERCIER, A., CORDEAU, J.-F. AND SOUMIS, F. (2005). A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers & Operations Research*, 32, 1451–1476.
- REXING, B., BARNHART, C. AND KNIKER, T. (2000). Airline fleet assignment with time windows. *Transportation Science*, 34, 1–20.
- RUSHMEIER, R.A. AND KONTOGIORGIS, S.A. (1997). Advances in the optimization of airline fleet assignment. *Transportation Science*, 31, 159–169.
- TALLURI, K.T. (1998). The four-day aircraft maintenance routing problem. *Transportation Science*, 32, 43–53.
- YU, G., ed. (1998). *Operations Research in the Airline Industry*. Boston: Kluwer.

CHAPITRE 6

A THEORETICAL COMPARISON OF FEASIBILITY CUTS FOR THE INTEGRATED AIRCRAFT ROUTING AND CREW PAIRING PROBLEM

Article écrit par Anne Mercier, soumis à *Transportation Science*, mars 2006.

L'approche principale proposée pour résoudre le problème intégré de la création d'itinéraires d'avions et d'horaires d'équipages consiste à résoudre, itérativement, un problème de rotations d'équipages modifié, où une coupe de réalisabilité est ajoutée au problème d'équipages jusqu'à ce que l'ensemble des connexions courtes utilisé par les équipages dans la solution courante soit réalisable pour les avions. Les coupes de réalisabilité peuvent être générées par une approche de décomposition de Benders, avec comme sous-problème le problème de routage d'avions, ou à partir d'une famille prédéfinie de coupes. Par exemple, COHN et BARNHART (2003) ont proposé une famille de coupes de réalisabilité interdisant un sous-ensemble de connexions courtes dit *minimalement non-réalisable* (MIS). Quel que soit le type de coupe choisi, on doit vérifier à l'aide d'un sous-problème, à chaque itération, si l'ensemble de connexions courtes utilisé par la solution courante d'équipages est réalisable pour les avions. On peut montrer que les rayons extrêmes du polyèdre dual de ce sous-problème n'ont pas de forme particulière. Les variables duales peuvent, en effet, prendre différentes valeurs dans un même rayon extrême. De plus, on peut montrer que les coupes de réalisabilité générées à partir d'une famille prédéfinie sont toutes incluses dans l'ensemble des coupes de Benders, et peuvent aussi être dominées par ces dernières.

Pour éviter de devoir identifier des rayons extrêmes, on peut borner le sous-problème dual. Pour ce faire, il suffit d'introduire des variables artificielles dans le sous-problème primal d'avions pour le rendre réalisable. On peut montrer que le choix de ces variables artificielles a un impact sur le polyèdre dual, et donc sur les coupes générées. En effet, l'ensemble des coupes générées à partir d'un sous-problème d'avions dans lequel est ajouté une seule variable artificielle contribuant à toutes les contraintes liantes de connexion courte correspond à l'ensemble des rayons extrêmes. Or, ce n'est pas le cas lorsqu'on introduit une variable artificielle différente pour chacune des contraintes liantes. Cette dernière stratégie peut effectivement favoriser certaines coupes dont les coefficients prennent tous la valeur un, même si ces coupes sont plus faibles. Par ailleurs, aucune des coupes de réalisabilité ne favorise l'inclusion des connexions courtes qui ne sont pas prises par la solution courante d'équipages, même si celles-ci sont incompatibles avec les connexions incluses dans la coupe.

Pour palier à cette faiblesse, on peut utiliser deux procédures de renforcement. Comme il est possible que le sous-problème primal de Benders soit dégénéré, il peut exister plus d'une solution au sous-problème dual à chaque itération de la méthode de décomposition de Benders. Même si chacune de ces solutions permet la construction d'une coupe valide, certaines peuvent mener à des coupes plus fortes que d'autres. Pour identifier une solution optimale au sous-problème dual qui mène à une coupe Pareto-optimale (une coupe qui n'est dominée par aucune autre), on doit résoudre un sous-problème dual auxiliaire après chaque résolution du sous-problème de Benders. La fonction-objectif du problème auxiliaire compare toutes les coupes de Benders possibles à un point de l'intérieur relatif de la région admissible du problème-maître. Dans ce cas, toutes les connexions courtes peuvent être incluses dans la fonction-objectif du sous-problème dual auxiliaire, même celles qui ne sont pas utilisées dans la solution courante d'équipages. On peut aussi utiliser une procédure de renforcement qui élimine l'écart restant dans les contraintes du sous-problème dual d'origine (non-borné), si possible. On peut montrer que ce nouveau problème est toujours réalisable

et borné, qu'il ne demande qu'un petit effort de calcul et qu'il peut être appliqué à tous les types de coupes de réalisabilité (pas seulement les coupes de Benders). En plus de favoriser les coupes incluant des connexions courtes qui ne sont pas prises par la solution courante d'équipages, cette procédure a aussi l'avantage de favoriser les coupes plus fortes, c'est-à-dire celles dont les coefficients ne sont pas tous égaux à un. On peut aussi montrer que si la procédure est appliquée à une coupe correspondant à un rayon extrême, alors la coupe renforcée correspondra elle aussi à un rayon extrême.

Il existe un type particulier de coupes de réalisabilité au problème intégré avions-équipages pour lesquelles on peut montrer que certaines combinaisons de rayons extrêmes donnent une coupe plus forte que celles générées directement à partir des rayons extrêmes. Ces coupes ont un membre de droite de valeur zéro, et elles interdisent l'utilisation des connexions courtes qui sont individuellement non-réalisables pour les avions. On peut en effet montrer qu'il existe une combinaison de rayons extrêmes dont la coupe correspondante interdit en même temps toutes les connexions courtes individuellement non-réalisables parmi celles qui sont prises dans la solution d'équipages courante, alors que les coupes correspondant à un rayon extrême n'interdisent, quant à elles, qu'une seule connexion courte non-réalisable à la fois.

A Theoretical Comparison of Feasibility Cuts
for the Integrated Aircraft Routing and
Crew Pairing Problem

ANNE MERCIER

École Polytechnique de Montréal

March 2005

Abstract

The integrated aircraft routing and crew pairing problem consists in determining a minimum-cost set of aircraft routes and crew pairings such that each flight leg is covered by one aircraft and one crew, and some side constraints are satisfied. Linking constraints impose minimum connection times for crews that depend on aircraft connections. The main solution approach for this problem consists in solving a constrained crew pairing problem iteratively, adding feasibility cuts until a solution is found where the short connection set used by the crew pairings is feasible for the aircraft routing problem. The feasibility cuts can be generated by a Benders decomposition approach in which aircraft routing is handled by the subproblem, or they can be selected from a predefined family. We perform a theoretical comparison of the different types of feasibility cuts. We also propose a simple procedure to strengthen these cuts. Computational experiments performed on test instances provided by two major airlines are presented to support the theoretical results.

Keywords: aircraft routing; crew pairing; integrated planning; Benders decomposition; extreme rays; feasibility cuts.

Introduction

The planning of airline operations is usually performed sequentially (see, *e.g.*, YU (1998)). The airline first solves a *flight scheduling problem* to determine each flight leg to be flown during a given period, with corresponding departure and arrival times. Then, the *fleet assignment* is performed to assign an aircraft type to each flight leg, taking into account the number of available aircraft of each type and the estimated profit. An *aircraft routing problem* is then solved, for each aircraft type, to determine a sequence of flight legs to be assigned to each individual aircraft so that each leg is flown exactly once and maintenance is performed at regular interval. The airline then creates minimum-cost crew pairings by solving a *crew scheduling problem* for each aircraft type, ensuring that every flight leg is covered and that all work rules are satisfied. A *pairing* is a sequence of duty periods separated by overnight rests, and a *duty period* is a sequence of flight legs separated by smaller rest periods, called *sits* or *connections*. Finally, by solving a *crew bidding problem* or a *crew rostering problem*, pairings are combined to form personalized monthly schedules which are assigned to employees. On the one hand, because the five steps are linked together, a sequential planning procedure is likely to yield suboptimal solutions. On the other hand, completely integrating all steps would yield an intractable model. This paper focuses on the integration of the aircraft routing and the crew scheduling problems.

Aircraft routing decisions have an impact on the set of feasible pairings since the minimum connection time required between two successive flight legs covered by the same crew depends on whether the same aircraft is used on both legs. CORDEAU *et al.* (2001), KLABJAN *et al.* (2002) and COHN and BARNHART (2003) have shown that integrating the two problems yields solutions that are significantly better than those obtained by solving the problems sequentially. A connection that is not long enough to be made by a crew when the two associated legs are flown by a

different aircraft is said to be *short*. CORDEAU *et al.* (2001) have introduced a model where one linking constraint per short connection is added to the aircraft routing and crew pairing formulations to ensure that a crew uses a short connection only if the two connecting flights are flown by the same aircraft. To handle these linking constraints, a solution approach based on Benders decomposition is used. The latter model was further enhanced by MERCIER *et al.* (2005) who have shown that reversing the order of the solution sequence, *i.e.*, solving the crew pairing problem in the Benders master problem as opposed to the aircraft routing problem, yields significant improvements over the approach of CORDEAU *et al.* (2001). Since the aircraft routing problem is usually modeled as a feasibility problem, all costs in the integrated model are associated with the crew pairings. Hence, when solving an aircraft routing subproblem, only feasibility information is transferred to the master problem (in the form of feasibility cuts). HUANG *et al.* (2003) have also proposed a Benders decomposition approach for the integrated problem, but instead of generating feasibility cuts from the dual subproblem variables, they generate them from primal information. COHN and BARNHART (2003) have proposed an integrated model which includes variables representing complete solutions to the aircraft routing problem instead of incorporating the aircraft routing formulation. The authors show that only a subset of the feasible solutions needs to be included in the model. They have also proposed an approach which first solves a crew pairing problem in which all short connections are allowed. If the set of short connections used in the solution to this crew pairing problem leads to a feasible aircraft routing problem, an optimal solution has been identified. Otherwise, a feasibility cut is introduced in the crew pairing problem to forbid a minimally infeasible subset of short connections, and the process is repeated. Finally, KLABJAN *et al.* (2002) have presented a partially integrated approach that solves a modified crew scheduling problem including additional constraints that count the number of available aircraft on the ground at any time.

The contribution of this paper is threefold. First, we present a theoretical comparison of different families of feasibility cuts that have been proposed in the literature for the integrated aircraft routing and crew scheduling problem. We discuss their strength and the procedures used for their generation. The discussion includes a comparison of Benders feasibility cuts and Benders optimality cuts through the description of the corresponding dual subproblem polyhedra. The latter cuts are generated from an aircraft routing primal subproblem that is made feasible by the introduction of artificial variables. In that case, the choice of artificial variables has an impact on the dual subproblem polyhedron, and thus, on the feasibility cuts being generated. Second, we show that in the case of cuts with a right-hand-side of value zero, linear combinations of extreme rays give stronger cuts than extreme rays. Third, we introduce a simple lifting procedure that can be applied to any type of feasibility cut for the integrated aircraft routing and crew scheduling problem. Computational experiments performed on test instances provided by two major airlines are presented to support the theoretical results.

The remainder of the article is organized as follows. The next section introduces some notation and a mathematical formulation of the problem. Section 6.2 presents the Benders feasibility cuts and a description of the dual subproblem polyhedron. Some families of predefined feasibility cuts are then introduced in Section 6.3. Section 6.4 compares different approaches to bound the Benders dual subproblem polyhedron and generate optimality cuts. Section 6.5 introduces two families of stronger cuts: Pareto-optimal cuts and feasibility cuts obtained from a simple lifting procedure. This is followed by computational experiments in Section 6.6, and by the conclusion.

6.1 Mathematical Formulation

We consider the *Extended Crew Pairing* (ECP) formulation proposed by COHN and BARNHART (2003) for the integrated aircraft routing and crew pairing problem. In this formulation, variables representing complete solutions to the maintenance routing problem are added to the classical crew pairing model. All maintenance constraints can thus be eliminated and a single aircraft routing solution is chosen through the use of a *convexity constraint*. COHN and BARNHART (2003) show that only a subset of the feasible solutions needs to be included in the model, *i.e.*, one column for each *unique and maximal maintenance-feasible short connection set* (UM). These columns can be generated individually and sequentially, in a preprocessing step, by solving a series of aircraft routing problems with additional constraints and a modified objective function. Although the aircraft routing problem is considered to be the easiest of the airline planning problems, the modified routing problem is a difficult combinatorial optimization problem (MERCIER *et al.* (2005)). Nevertheless, we consider the ECP formulation in this paper to facilitate the comparison between the different feasibility cuts that have been proposed in the literature. Our conclusions should also apply to alternative formulations such as that of MERCIER *et al.* (2005). In addition, once the UM columns have all been generated, the ECP formulation can be used to perform sensitivity analysis on the crew pairing problem.

Let L be the set of flight legs, K the set of feasible crew pairings, R the set of feasible aircraft routing solutions, and S the set of short connections included in at least one aircraft routing solution. Denote by c_k the cost of pairing $k \in K$. For every leg $i \in L$ and every pairing $k \in K$, define a binary constant a_{ki} that takes value 1 if leg $i \in L$ is covered by pairing k . For every short connection $s \in S$ and every pairing $k \in K$ (resp. routing $r \in R$), define a binary constant d_{ks} (resp. b_{rs}) that takes value 1 if connection $s \in S$ is included in pairing $k \in K$ (resp. routing $r \in R$). Finally, let x_k

and y_r be binary variables that take value 1 if and only if pairing $k \in K$ and routing $r \in R$ is used in the solution, respectively. The ECP model can be stated as follows:

$$\text{Minimize } \sum_{k \in K} c_k x_k \quad (6.1)$$

subject to

$$\sum_{k \in K} a_{ki} x_k = 1 \quad (i \in L) \quad (6.2)$$

$$\sum_{r \in R} b_{rs} y_r - \sum_{k \in K} d_{ks} x_k \geq 0 \quad (s \in S) \quad (6.3)$$

$$\sum_{r \in R} y_r = 1 \quad (6.4)$$

$$x_k \in \{0, 1\} \quad (k \in K) \quad (6.5)$$

$$y_r \in \{0, 1\} \quad (r \in R). \quad (6.6)$$

The objective function (6.1) minimizes crew pairing costs. No costs are associated with aircraft routings. Constraints (6.2) together with constraints (6.5) ensure that each leg is covered by exactly one crew. Constraints (6.3) guarantee that a crew does not change aircraft when the connection time is too small. Constraint (6.4) together with constraints (6.6) ensure that exactly one solution to the aircraft routing problem is chosen.

6.2 Benders Decomposition

The ECP model (6.1)-(6.6) includes both crew pairing and aircraft routing variables. Benders decomposition (see, *e.g.*, BENDERS (1962); NEMHAUSER and WOLSEY (1988)) can be used to reformulate the problem so as to reduce the number of

variables at the expense of an increase in the number of constraints. The additional constraints can, however, be generated dynamically only when they are violated. In most applications, only a very small subset of constraints needs to be generated for an optimal solution to be identified.

Let $\mathbf{X}$ be the set of solutions satisfying the crew constraints (6.2) and (6.5). For a given vector $\bar{\mathbf{x}} \in \mathbf{X}$, the ECP reduces to the following primal subproblem (PSP) involving only aircraft variables:

$$\text{Minimize } 0 \tag{6.7}$$

subject to

$$\sum_{r \in R} b_{rs} y_r \geq \sum_{k \in K} d_{ks} \bar{x}_k \quad (s \in S) \tag{6.8}$$

$$\sum_{r \in R} y_r = 1 \tag{6.9}$$

$$y_r \geq 0 \quad (r \in R). \tag{6.10}$$

Observe that upper bounds on the y_r variables are not needed because of constraints (6.9) and that the integrality constraints on the aircraft variables y_r have been replaced with non-negativity constraints. In fact, once the crew variables x_k are given binary values, the polyhedron corresponding to (6.8)-(6.10) has integer extreme points (COHN and BARNHART (2003)). Let $\boldsymbol{\alpha} = (\alpha_s \geq 0 \mid s \in S)$ and β be the dual variables associated with constraints (6.8) and (6.9), respectively. When convexity constraint (6.9) is multiplied by -1 , the dual of (6.7)-(6.10) is the following dual subproblem (DSP):

$$\text{Maximize } \sum_{k \in K} \sum_{s \in S} d_{ks} \bar{x}_k \alpha_s - \beta \tag{6.11}$$

subject to

$$\sum_{s \in S} b_{rs} \alpha_s - \beta \leq 0 \quad (r \in R) \quad (6.12)$$

$$\alpha_s \geq 0 \quad (s \in S). \quad (6.13)$$

Since $b_{rs} \geq 0$, for all $r \in R$ and $s \in S$, one can observe that $\beta \geq 0$ by (6.12) and (6.13). The dual subproblem is always feasible since the null vector $\mathbf{0}$ satisfies constraints (6.12)-(6.13). Hence, the primal subproblem is either infeasible or feasible and bounded. For the dual subproblem to be bounded (and the primal subproblem feasible), the value of its objective function (6.11) must be non-positive for every extreme ray of its feasible region. Let Δ denote the polyhedron defined by constraints (6.12)-(6.13). One can see that Δ is a pointed polyhedral cone, *i.e.*, it has a unique extreme point, the null vector, and a multitude of extreme rays. Let R_Δ be the set of extreme rays of Δ . One can notice that only the objective function of the dual subproblem (6.11)-(6.13) contains crew information. This implies that the feasible region of the dual subproblem is independent of the crew pairing solution and that all the extreme rays of Δ could be enumerated *a priori*. Model (6.1)-(6.6) can thus be reformulated as the following Benders master problem (MP):

$$\text{Minimize } \sum_{k \in K} c_k x_k \quad (6.14)$$

subject to

$$\sum_{k \in K} a_{ki} x_k = 1 \quad (i \in L) \quad (6.15)$$

$$\sum_{k \in K} \sum_{s \in S} d_{ks} \alpha_s x_k \leq \beta \quad ((\alpha, \beta) \in R_\Delta) \quad (6.16)$$

$$x_k \in \{0, 1\} \quad (k \in K). \quad (6.17)$$

The Benders master problem is comprised of the crew pairing problem (constraints (6.14), (6.15) and (6.17)) and of the set of feasibility constraints (6.16) which ensure that the values given to the crew pairing variables x_k lead to a bounded dual subproblem (a feasible primal aircraft routing subproblem). In general, model (6.14)-(6.17) contains more constraints than model (6.1)-(6.6) but most feasibility constraints are inactive in an optimal solution. Hence, these constraints need not be enumerated exhaustively but can instead be generated dynamically by iterating between a relaxed master problem and the subproblem. The relaxed master problem contains constraints (6.15) and (6.17) as well as a subset of the Benders cuts (6.16). The optimal solution of the relaxed Benders master problem is used to set up constraints (6.8) in the primal subproblem at every iteration. If the primal subproblem is feasible, an optimal solution to the ECP model has been obtained. Otherwise, an extreme ray of R_Δ violating one of the constraints (6.16) is identified. Hence, exactly one constraint is added to the relaxed Benders master problem at each iteration. For further notational convenience, one can define $x_s = \sum_{k \in K} d_{ks} x_k$.

6.2.1 Extreme rays

Assume that $q = (\alpha^q, \beta^q)$ is a ray of Δ . We now define the following additional notation:

- $S^q = \{s \in S \mid \alpha_s^q > 0\}$: the set of short connections for which the corresponding dual variable is positive in q .
- $R^q = \{r \in R \mid \sum_{s \in S^q} b_{rs} \alpha_s^q = \beta^q\}$: the set of aircraft routings for which the corresponding constraint (6.12) is active at q .
- $n(R^q)$: the maximum number of linearly independent elements in R^q .

- $m^q = \max_{r \in R^q} \sum_{s \in S^q} b_{rs}$: the maximum number of short connections from S^q in an aircraft routing from R^q (maximum cardinality).
- $\hat{R}^q = \{r \in R^q \mid \sum_{s \in S^q} b_{rs} = m^q\}$: the set of maximum cardinality aircraft routings from R^q .

In the remainder of the paper, we will refer interchangeably to a routing $r \in R$ and to the constraint (6.12) associated with it. The same is true for a short connection $s \in S$ and its corresponding dual variable α_s .

Proposition 1. *A ray q is an extreme ray of Δ if and only if $n(R^q) = |S^q|$.*

Proof. Recall that a ray t of a polyhedral cone $C \subseteq R^n$ is an extreme ray if and only if there are $n - 1$ linearly independent constraints that are active at t (see e.g. BERTSIMAS and TSITSIKLIS (1997)). In the Benders DSP, the total number of variables is $|S| + 1$. For q to be an extreme ray of Δ , the maximum number of linearly independent constraints satisfied at equality by q thus has to be equal to $|S|$. Because of constraints (6.13), the variables taking the value 0 in q correspond to active non-negativity constraints. Therefore, only the number of positive variables α_s , i.e., $|S^q|$, has to be considered and compared to the maximum number of linearly independent constraints (6.12) which are active at q , i.e., $n(R^q)$. $\square$

Let $a(q)$ be a ray where $\alpha_s^{a(q)} = 1$, for all $s \in S^q$, and $\beta^{a(q)} = m^q$.

Proposition 2. *If q is an extreme ray of Δ and $R^q = \hat{R}^q$, i.e., all active constraints are of maximum cardinality, then there exists $\nu > 0$ such that $q = \nu a(q)$.*

Proof. The system of equations obtained by replacing each inequality $r \in R^q$ with the equality $\sum_{s \in S^q} b_{rs} \alpha_s^q = \beta^q$ is equivalent to the original system. Consider a system of

equations having a unique solution (q is an extreme ray of Δ), where the coefficient of each variable in each constraint is equal to 1 and the right-hand side of all constraints has the same value (both conditions are satisfied here). If all constraints contain the same number of positive variables ($R^q = \hat{R}^q$), then all variables must have the same value in the solution and the right-hand-side of each constraint is equal to the number of positive variables multiplied by their common value. Therefore, $\exists \nu > 0$ such that $q = \nu a(q)$. $\square$

Other conditions on the extreme rays of Δ can be stated, yielding various specific forms for the extreme rays. However, the following simple example shows that the extreme rays can take several forms and stating all of them would be pointless.

Example 1. *A possible set of constraints for the Benders DSP:*

$$\alpha_A + \alpha_B + \alpha_C \leq \beta \quad (6.18)$$

$$\alpha_A + \alpha_B + \alpha_D \leq \beta \quad (6.19)$$

$$\alpha_A + \alpha_C + \alpha_D \leq \beta \quad (6.20)$$

$$\alpha_B + \alpha_C + \alpha_D \leq \beta \quad (6.21)$$

$$\alpha_B + \alpha_D + \alpha_E \leq \beta \quad (6.22)$$

$$\alpha_A + \alpha_E \leq \beta \quad (6.23)$$

$$\alpha_C + \alpha_F \leq \beta \quad (6.24)$$

$$+ \alpha_G \leq \beta \quad (6.25)$$

$$\alpha_A, \alpha_B, \alpha_C, \alpha_D, \alpha_E, \alpha_F, \alpha_G \geq 0. \quad (6.26)$$

Let Δ_1 denote the polyhedron defined by constraints (6.18)-(6.26) and let R_{Δ_1} be the set of extreme rays of Δ_1 . Consider the following four rays of Δ_1 :

- $a = \{1, 1, 0, 0, 0, 0, 0, 2\} \implies S^a = \{A, B\}$, $R^a = \{(6.18), (6.19)\}$ and $m^a = 2$.
One can observe that $n(R^a) = 1$ since, when the null variables α_C and α_D are removed, constraints (6.18) and (6.19) are linearly dependent. The ray a is thus not an extreme ray of Δ_1 since $|S^a| > n(R^a)$.
- $b = \{1, 1, 0, 0, 1, 0, 0, 2\} \implies S^b = \{A, B, E\}$, $R^b = \{(6.18), (6.19), (6.22), (6.23)\}$, $m^b = 2$, $n(R^b) = 3$ and $\hat{R}^b = \{(6.18), (6.19), (6.22), (6.23)\}$. Since $|S^b| = n(R^b)$, b is an extreme ray of Δ_1 . Here, $\hat{R}^b = R^b$ and thus $\alpha_A^b = \alpha_B^b = \alpha_E^b = 1$ and $\beta^e = m^b$, as in Proposition 2.
- $c = \{1, 1, 0, 0, 1, 2, 0, 2\} \implies S^c = \{A, B, E, F\}$, $R^c = \{(6.18), (6.19), (6.22), (6.23), (6.24)\}$, $m^c = 2$, $n(R^c) = 4$ and $\hat{R}^c = \{(6.18), (6.19), (6.22), (6.23)\} \neq R^c$. Since $|S^c| = n(R^c)$, c is an extreme ray of Δ_1 . Here, $\alpha_A^c = \alpha_B^c = \alpha_E^c = 1$ and $\beta^c = m^c$, but $\alpha_F^c = 2$.
- $d = \{2, 1, 1, 1, 2, 0, 0, 4\} \implies S^d = \{A, B, C, D, E\}$, $R^d = \{(6.18), (6.19), (6.20), (6.22), (6.23)\}$, $m^d = 3$, $n(R^d) = 5$ and $\hat{R}^d = \{(6.18), (6.19), (6.20), (6.22)\} \neq R^d$. Since $|S^d| = n(R^d)$, d is an extreme ray of Δ_1 . Here, $\alpha_A^d = \alpha_E^d = 1$, but $\beta^d \neq m^d$ and $\alpha_B^d = \alpha_C^d = \alpha_D^d = 2$.

6.3 Some Predefined Families of Feasibility Cuts

One can solve the integrated problem iteratively, without using the Benders reformulation, by first solving a crew pairing problem in which all short connections are allowed. If the aircraft maintenance routing problem is feasible with the chosen set of short connections, then an optimal solution to the integrated problem has been identified. Otherwise, a feasibility cut is added to the crew pairing problem and the process continues until the solution to the crew pairing problem uses a set of short connections for which the aircraft routing problem is feasible.

This solution process is called the *Constrained Crew Pairing* (CCP) by COHN and BARNHART (2003), who have proposed three families of feasibility cuts. The first family simply forbids the current crew pairing solution:

$$\sum_{k \in K^t} x_k \leq |K^t| - 1, \quad (6.27)$$

where $K^t = \{k \in K \mid x_k^t = 1\}$ and $\mathbf{x}^t$ is the optimal crew pairing solution at iteration t . The authors remark, however, that this type of cut is not very efficient since it prohibits a maintenance infeasible set of pairings while there may exist other sets of pairings using the same set of short connections. Hence, the second type of feasibility cut that is proposed prohibits the short connection set F^t used by the optimal solution at iteration t :

$$\sum_{k \in K} \sum_{s \in F^t} d_{ks} x_k \leq |F^t| - 1. \quad (6.28)$$

The authors observe that this type of cut may also be inefficient if a subset of incompatible short connections from F^t is attractive for the crew pairing problem and is likely to be chosen in several successive iterations. For example, if $F^t = \{A, B, C, D, E\}$, where A , B and C are incompatible, the next iterations could yield solutions with short connection sets $\{A, B, C, D\}$, $\{A, B, C, E\}$ and $\{A, B, C\}$, which are all maintenance infeasible.

To circumvent this weakness, the authors finally propose a family of cuts that directly prohibits a *Minimally Infeasible Subset* (MIS) $\tilde{F}^t$ of F^t . An MIS is an infeasible set of short connections such that the removal of any element from the set yields a feasible

subset. The resulting cut is the following:

$$\sum_{k \in K} \sum_{s \in \tilde{F}^t} d_{ks} x_k \leq |\tilde{F}^t| - 1. \quad (6.29)$$

To determine a set $\tilde{F}^t$, one may solve the following integer problem denoted by (PMIS):

$$\text{Minimize } \sum_{s \in F^t} f_s \quad (6.30)$$

subject to

$$\sum_{s \in F^t \setminus S(r)} f_s \geq 1 \quad (r \in R) \quad (6.31)$$

$$f_s \in \{0, 1\} \quad (s \in F^t), \quad (6.32)$$

where f_s is a binary variable indicating whether the short connection $s \in F^t$ is included in $\tilde{F}^t$ and $S(r)$ is the set of short connections used in routing $r \in R$. Constraints (6.31) require that for every feasible aircraft solution $r \in R$, there be at least one element of $\tilde{F}^t$ that is not in $S(r)$. The latter constraints ensure that the chosen subset $\tilde{F}^t$ is maintenance infeasible. The objective function (6.30) finds the smallest maintenance infeasible subset of F^t .

Proposition 3. *All MIS feasibility cuts (6.29) correspond to an extreme ray of Δ .*

Proof. Let $c = (\alpha^c, \beta^c)$ be a point corresponding to an MIS cut formed from the short connection set $\tilde{F}^t$, where $\alpha_s^c = 1, \forall s \in \tilde{F}^t$, and $\beta^c = |\tilde{F}^t| - 1$. Since no routing contains more than $|\tilde{F}^t| - 1$ short connections from $\tilde{F}^t$ ($\tilde{F}^t$ is a minimally infeasible subset) and $\alpha_s^c = 0, \forall s \in S \setminus \tilde{F}^t$, each constraint (6.12) contains at most $|\tilde{F}^t| - 1$ positive variables (of value 1) and is satisfied. The point c is thus feasible since the constraints (6.13) are satisfied from the definition of c . One can also observe that νc is in $\Delta, \forall \nu > 0$, and

that c is thus a ray of Δ . Since there exists a feasible aircraft routing for each subset of $\tilde{F}^t$, there exists a feasible routing for each of the $|\tilde{F}^t|$ subsets of size $|\tilde{F}^t| - 1$. Those distinct subsets (and only those) each correspond to a constraint in the Benders dual subproblem that is active at c , since $\beta^c = |\tilde{F}^t| - 1$ and $\alpha_s^c = 1, \forall s \in \tilde{F}^t$. There are thus exactly $|\tilde{F}^t|$ linearly independent active constraints at c . Hence, $n(R^c) = |\tilde{F}^t| = |S^c|$ and c is an extreme ray of Δ . $\square$

Proposition 4. *Not all extreme rays of Δ correspond to a MIS cut.*

Proof. In Example 1, the extreme ray $c = \{1, 1, 0, 0, 1, 2, 0, 2\}$ does not correspond to an MIS. Recall that $x_s = \sum_{k \in K} d_{ks} x_k$. One can observe that c gives a stronger feasibility cut ($x_A + x_B + x_E + 2x_F \leq 2$) than the extreme ray $b = \{1, 1, 0, 0, 1, 0, 0, 2\}$ which corresponds to an MIS ($x_A + x_B + x_E \leq 2$). $\square$

Furthermore, the following simple example shows an extreme ray of Δ that does not correspond to an MIS even if, in this case, $\alpha_s = 1, \forall s \in S$, and $S = F^t$.

Example 2. *A possible set of constraints for the Benders DSP:*

$$\alpha_A \leq \beta \tag{6.33}$$

$$\alpha_B \leq \beta \tag{6.34}$$

$$\alpha_C \leq \beta \tag{6.35}$$

$$\alpha_A, \alpha_B, \alpha_C \geq 0. \tag{6.36}$$

If the current crew pairing solution uses all short connections, i.e., $F^t = \{A, B, C\}$, the aircraft routing problem is infeasible and there are three possible MIS: $\{A, B\}$, $\{A, C\}$ and $\{B, C\}$. For instance, if $\tilde{F}^t = \{A, B\}$, then the MIS cut is: $x_A + x_B \leq 1$. There are four non-trivial extreme rays, of which three correspond to the three MIS.

The last one, $\{1, 1, 1, 1\}$ does not correspond to an MIS and it gives the strongest cut: $x_A + x_B + x_C \leq 1$.

Proposition 5. *An MIS feasibility cut (6.29) generated at iteration t can be lifted by adding to the corresponding MIS a short connection that is incompatible with all feasible subsets in $\tilde{F}^t$. If such a short connection exists, the resulting cut is stronger and it corresponds to an extreme ray of Δ .*

Proof. From Proposition 3, a ray q with $\alpha_s = 1, \forall s \in \tilde{F}^t$, and $\beta = |\tilde{F}^t| - 1$ is an extreme ray of Δ corresponding to the MIS $\tilde{F}^t$. Recall that $n(R^q)$ is the maximum number of linearly independent active dual constraints (6.12) at q , and $S^q = \tilde{F}^t$ is the set of positive variables in q . Since q is an extreme ray, $n(R^q) = |S^q|$. If there exists a short connection $s_1 \in S \setminus S^q$ which is incompatible with all feasible subsets of $\tilde{F}^t$, then a new ray $g = (\alpha^g, \beta^g)$ is obtained, where $\alpha_{s_1} \geq 1$, $\alpha_s = 1, \forall s \in S^q$, and $\beta^g = \beta^q = |\tilde{F}^t| - 1$. If s_1 exists, then there exists a feasible routing using s_1 while using a strictly smaller number of short connections from $\tilde{F}^t$ than the value of β^g . Therefore, $\alpha_{s_1}^g \geq 1$ and the cut corresponding to g is valid and stronger than the one corresponding to q . In addition, it is always possible to choose for α_{s_1} a value for which a new dual constraints (6.12) is active, without modifying any other value. When $\alpha_{s_1}^g = \min_{r \in R \setminus R^q} \beta^g - \sum_{s \in S^q} b_{rs}$, a new dual constraints (6.12) is satisfied at equality and $n(R^g) = n(R^q) + 1 = |S^q| + 1 = |S^g|$. The ray g is thus an extreme ray of Δ . $\square$

The lifted MIS cut can be stated as follows:

$$\sum_{k \in K} \sum_{s \in \tilde{F}^t} d_{ks} x_k + \sum_{k \in K} l_{s_1} d_{ks_1} x_k \leq |\tilde{F}^t| - 1, \quad (6.37)$$

where s_1 is a short connection $s \in S \setminus \tilde{F}^t$ that is incompatible with all feasible subsets of $\tilde{F}^t$ and l_{s_1} is the value taken by α_{s_1} in the corresponding extreme ray. An MIS cut

can be lifted repeatedly, *i.e.*, a set of more than one incompatible short connections can be added to the cut sequentially, as long as the added short connections are also incompatible with each other.

Recall that the extreme ray $b = \{1, 1, 0, 0, 1, 0, 0, 2\}$ from Example 1 corresponds to the MIS $\{A, B, E\}$ and that $n(R^b) = 3$. When looking at the dual constraints, one can see that the short connection F is incompatible with all three feasible subsets $\{A, B\}$, $\{A, E\}$, and $\{B, E\}$. The cut $x_A + x_B + x_E + x_F \leq 2$ is thus a lifting of the MIS cut $x_A + x_B + x_E \leq 2$ and it corresponds to the ray $e = \{1, 1, 0, 0, 1, 1, 0, 2\}$. However, the lifted cut does not correspond to an extreme ray of Δ_1 since $n(R^e) < |S^e|$ (there is one more positive dual variable than in b , but no new active constraint). Nonetheless, one can see that the dual constraint (6.23) becomes active when $\alpha_F = 2$. In fact, the lifted cut $x_A + x_B + x_E + 2x_F \leq 2$ is stronger and it corresponds to the extreme ray $c = \{1, 1, 0, 0, 1, 2, 0, 2\}$.

6.4 Bounding the Benders Dual Subproblem

When artificial variables are added to constraints (6.8) or (6.9) to ensure the feasibility of the Benders PSP, the Benders DSP becomes bounded and Benders feasibility cuts (6.16) are no longer needed in the Benders MP. However, because a cost is associated with each artificial variable, the PSP becomes an optimization problem and Benders optimality cuts must then be considered in the MP. This type of formulation is interesting since it is generally computationally easier to generate optimality cuts than feasibility cuts. Indeed, the latter requires the identification of a non-basic dual variable which causes unboundedness. Depending on the LP solver being used, this may be a difficult task, especially if it is the primal form of the subproblem which is solved. There are three different ways of making the primal subproblem feasible:

one can either introduce (a) one artificial variable in each short connection linking constraint (6.8), (b) a unique artificial variable appearing in every constraint (6.8), or (c) a unique artificial variable appearing in the convexity constraint (6.9). Since this choice has an impact on the dual subproblem polyhedron, all three models are examined below.

6.4.1 Model (a): One artificial variable in each short connection linking constraint

Recall that $\mathbf{X}$ is the set of solutions satisfying the crew constraints (6.2) and (6.5). For a given vector $\bar{\mathbf{x}} \in \mathbf{X}$, when model (a) is chosen, ECP reduces to the following feasible primal subproblem (FPSPa) involving only aircraft variables:

$$\text{Minimize } \sum_{s \in S} c_s a_s \quad (6.38)$$

subject to

$$\sum_{r \in R} b_{rs} y_r + a_s \geq \sum_{k \in K} d_{ks} \bar{x}_k \quad (s \in S) \quad (6.39)$$

$$-\sum_{r \in R} y_r = -1 \quad (6.40)$$

$$y_r \geq 0 \quad (r \in R) \quad (6.41)$$

$$a_s \geq 0 \quad (s \in S), \quad (6.42)$$

where a_s is an artificial variable associated with short connection $s \in S$ and c_s is the cost of using the artificial variable a_s . As is common in the first phase of the simplex algorithm, all artificial variables can be assumed to have an equal cost, $\bar{c}$. We can also assume, w.l.o.g., that $\bar{c} = 1$ since no other cost is present in FPSPa.

Recall that F^t is the short connection set used by the optimal crew pairing solution at iteration t , *i.e.*, $F^t = \{s \in S \mid \sum_{k \in K} d_{ks} \bar{x}_k = 1\}$. The optimal solution of FPSPa chooses an aircraft routing solution that minimizes the sum of the artificial variables, and thus maximizes the number of short connections used among those in F^t . Recall that $S(r)$ is the set of short connections used in routing r , and let $\bar{r}^t$ be the FPSPa aircraft routing solution at iteration t . HUANG *et al.* (2003) have proposed a family of cuts generated from $S(\bar{r}^t)$ as follows :

$$\sum_{k \in K} \sum_{s \in F^t} d_{ks} x_k \leq |S(\bar{r}^t) \cap F^t|. \quad (6.43)$$

Since $|S(\bar{r}^t) \cap F^t|$ is the maximum number of short connections that a feasible maintenance aircraft routing solution can use among those chosen by the current crew pairing solution, PSP is infeasible when $|S(\bar{r}^t) \cap F^t| < |F^t|$, and cuts (6.43) are valid feasibility cuts. One can notice that these cuts are stronger than feasibility cuts (6.28), but, as for MIS cuts (6.29), they can never include a short connection $s \in S \setminus F^t$.

Proposition 6. *When F^t is an MIS, cuts (6.43) are identical to the MIS cuts (6.29).*

Proof. If $\tilde{F}^t = F^t$, then $|S(\bar{r}^t) \cap F^t| = |\tilde{F}^t| - 1$. □

Proposition 7. *Feasibility cuts (6.43) do not necessarily correspond to an extreme ray of Δ .*

Proof. If $F^t = \{A, B, E, F\}$ in Example 1, then cut (6.43) is $x_A + x_B + x_E + x_F \leq 2$. One can easily see that the corresponding ray $r = \{1, 1, 0, 0, 1, 1, 0, 2\}$ is not an extreme ray of Δ_1 since $|S^r| = 4 \neq n(R^r) = 3$. □

The dual of (6.38)-(6.42) is the following bounded dual subproblem (BDSPa):

$$\text{Maximize } \sum_{k \in K} \sum_{s \in S} d_{ks} \bar{x}_k \alpha_s - \beta \quad (6.44)$$

subject to

$$\sum_{s \in S} b_{rs} \alpha_s - \beta \leq 0 \quad (r \in R) \quad (6.45)$$

$$\alpha_s \leq c_s \quad (s \in S) \quad (6.46)$$

$$\alpha_s \geq 0 \quad (s \in S). \quad (6.47)$$

No new dual variable is introduced, but the value of each α_s variable is now bounded by the cost of the associated artificial variable. Let Δ^{B_a} denote the polyhedron defined by constraints (6.45)-(6.47) and let $P_{\Delta^{B_a}}$ be the set of extreme points of Δ^{B_a} . Introducing the additional free variable z_0 , the integrated problem can thus be reformulated as the following Benders master problem (MP2a):

$$\text{Minimize } \sum_{k \in K} c_k x_k + z_0 \quad (6.48)$$

subject to

$$\sum_{k \in K} a_{ki} x_k = 1 \quad (i \in L) \quad (6.49)$$

$$-z_0 + \sum_{k \in K} \sum_{s \in S} d_{ks} \alpha_s x_k \leq \beta \quad ((\alpha, \beta) \in P_{\Delta^{B_a}}) \quad (6.50)$$

$$x_k \in \{0, 1\} \quad (k \in K). \quad (6.51)$$

The value of z_0 is restricted to be larger than or equal to the optimal value of the dual subproblem, associated with an extreme point, by optimality constraints (6.50).

One can notice that the difference between Benders feasibility cuts (6.16) and Benders optimality cuts (6.50) is the value of the cost associated with the artificial variables. This value is minimized by the Benders master problem MP2a and is equal to zero at optimality. Benders optimality cuts, in this problem, are in fact feasibility cuts.

Recall that R^q is the set of aircraft routings for which the corresponding dual constraint (6.12) in DSP is active at ray q and $n(R^q)$ is the maximum number of linearly independent elements in R^q . Now, let H^q be the complete set of active constraints at q (without the non-negativity constraints) and $n(H^q)$ be the maximum number of linearly independent elements in H^q . For example, in BDSPa, only the active dual constraints (6.45) which are active at q are included in R^q , and H^q is comprised of all active constraints (6.45) and (6.46).

Proposition 8. *A point p is an extreme point of Δ^{B_a} if and only if $n(H^p) \geq |S^p| + 1$.*

Proof. Recall that a point p of a polyhedron P of dimension n is an extreme point if and only if there are at least n linearly independent active constraints of P at p . In the Benders BDSPa, the total number of variables is $|S| + 1$ (with variable β). For p to be an extreme point of Δ^{B_a} , the maximum number of linearly independent constraints satisfied at equality by p thus has to be greater than or equal to $|S| + 1$. Since a non-negativity constraint is active for every null variable α , one does not need to consider constraints (6.47) and variables $\alpha_s, s \in S \setminus S^p$. Therefore, the point p is an extreme point of Δ^{B_a} if and only if $n(H^p) \geq |S^p| + 1$. $\square$

Proposition 9. *If p is an extreme point of Δ^{B_a} , then $n(S^p) \leq |S^p|$, i.e., the maximum number of linearly independent active constraints (6.45) at p is bounded by the number of positive variables α at the point.*

Proof. There are a total of $|S^p| + 1$ positive variables contributing to constraints (6.45). The maximum number of linearly independent active constraints (6.45) at p ,

$n(S^p)$, is thus bounded by $|S^p| + 1$. Furthermore, since the variable β is present, with the same coefficient, in all constraints (6.45), the latter bound can be improved, and $n(S^p) \leq |S^p|$. $\square$

Proposition 10. *If p is an extreme point of Δ^{B_a} , then $\alpha_s^p = \bar{c}$ for at least one short connection $s \in S^p$, and $\beta^p = \max_{r \in R} \sum_{s \in S^p} b_{rs} \alpha_s^p$.*

Proof. If $p \in P_{\Delta^{B_a}}$, then $n(H^p) \geq |S^p| + 1$ in BDSPa (see Proposition 8). Since there cannot be more than $|S^p|$ linearly independent active constraints (6.45) (see Proposition 9), or more than $|S^p|$ active constraints (6.46), at least one constraint of each type is satisfied at equality at p , i.e., $\alpha_s = \bar{c}$ for at least one short connection $s \in S^p$, and $\beta^p = \max_{r \in R} \sum_{s \in S^p} b_{rs} \alpha_s^p$. $\square$

Proposition 11. *All extreme rays of Δ correspond to an extreme point of Δ^{B_a} .*

Proof. If $q = (\alpha^q, \beta^q) \in R_{\Delta}$, then, $n(R^q) = |S^q|$ in DSP (see Proposition 1). By multiplying every element of q by a positive constant, one can rescale the extreme ray such that at least one of the constraints (6.46) is active at q and all the others are satisfied. Since the rescaling does not modify R^q or S^q , and the set of constraints (6.45) is the same as the set (6.12), then $n(R^q) = |S^q|$ also in BDSPa. Hence, $n(H^q) \geq |S^q| + 1$ in BDSPa and, from Proposition 8, $q \in P_{\Delta^{B_a}}$. $\square$

Proposition 12. *Not all extreme points of Δ^{B_a} correspond to an extreme ray of Δ .*

Proof. The null vector is the unique extreme point of Δ . All other feasible solutions to DSP are rays of Δ . Since all extreme points of Δ^{B_a} are feasible solutions to Δ , they thus correspond to a ray of Δ . However, not all extreme points of Δ^{B_a} correspond to an extreme ray of Δ . One need only consider the point $a = \{1, 1, 0, 0, 0, 0, 2\}$ in Example 1. The point a is an extreme point of $\Delta_1^{B_a}$ since $n(H^a) = 3 \geq |S^a| + 1 = 3$. However, a is not an extreme ray of Δ since $n(R^a) = 1 < |S^a| = 2$. $\square$

Since $F^t = \{s \in S \mid \sum_{k \in K} d_{ks} \bar{x}_k = 1\}$, let $v_a(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p$ be the value of the objective function (6.44) of BDSPa at the point p for a given set F^t of short connections, and let $v_a^*(F^t)$ be the optimal value of (6.44) for the same problem. Let $\bar{P}_{\Delta^{B_a}}^{F^t}$ be the set of extreme points of Δ^{B_a} that maximize the value of (6.44) for a given set F^t , i.e., $\bar{P}_{\Delta^{B_a}}^{F^t} = \{p \in P_{\Delta^{B_a}} \mid v_a(p, F^t) = v_a^*(F^t)\}$.

Proposition 13. *For a given set of short connections F^t , not all optimal extreme points of Δ^{B_a} correspond to an extreme ray of Δ .*

Proof. If $F^t = \{A, B, C, D, E, F, G\}$, in Example 1, and $\bar{c} = 1$, then $v_a^*(F^t) = 4$ and $f = \{1, 1, 1, 1, 1, 1, 3\} \in \bar{P}_{\Delta_1}^{F^t}$. One can easily see that f , an extreme point of $\Delta_1^{B_a}$, is not an extreme ray of Δ_1 since $|S^f| = 7 \neq n(R^f) = 5$. $\square$

Proposition 14. *If, for a given set F^t , p is an optimal extreme point of Δ^{B_a} , then there exists $r_i \in R^p$ such that $\sum_{s \in F^t} b_{r_i s} \alpha_s^p = \beta^p$, i.e., there exists at least one active dual constraint (6.45) in BDSPa in which all contributions come from the variables α from F^t ($b_{r_i s} = 0$, for all $s \in (S^p \setminus F^t)$).*

Proof. If, for a given set F^t , $p \in \bar{P}_{\Delta^{B_a}}^{F^t}$, then there is at least one constraint (6.45) which is active at p (see Proposition 10). Hence, $R^p \neq 0$ and $\sum_{s \in S} b_{r s} \alpha_s^p = \beta^p$, $\forall r \in R^p$. Recall that, for a given set F^t , the value of the objective function (6.44) of BDSPa at the point p is $v_a(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p$. Let $r_1 \in R^p$ be any active constraint (6.45) at p . Therefore, $\beta^p = \sum_{s \in S} b_{r_1 s} \alpha_s^p$ and $v_a(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \sum_{s \in S} b_{r_1 s} \alpha_s^p$. One can observe that the variables from $S \setminus F^t$ can never positively contribute to the objective function value. If $\sum_{s \in F^t} b_{r s} \alpha_s^p < \beta^p$, $\forall r \in R^p$, i.e., all active constraints (6.45) include contributions from variables α in $S \setminus F^t$, then one can construct a point $p' = (\alpha^{p'}, \beta^{p'})$, where (i) $\alpha_s^{p'} = \alpha_s^p$, $\forall s \in F^t$, (ii) $\alpha_s^{p'} = 0$, $\forall s \in S \setminus F^t$, and (iii) $\beta_{p'} = \beta_p - \min_{r \in R^p} \sum_{s \in S \setminus F^t} b_{r s} \alpha_s^p$. One can observe that p' is feasible since it satisfies all constraints (6.45)-(6.47) from BDSPa. The objective value at p' is

$v_a(p', F^t) = \sum_{s \in F^t} \alpha_s^{p'} - \beta^{p'} = \sum_{s \in F^t} \alpha_s^p - \beta_p + \min_{r \in R^{p'}} \sum_{s \in S \setminus F^t} b_{rs} \alpha_s^p = v_a(p, F^t) + \min_{r \in R^{p'}} \sum_{s \in S \setminus F^t} b_{rs} \alpha_s^p$. Since $\sum_{s \in F^t} b_{rs} \alpha_s^p < \beta^p, \forall r \in R^p, \min_{r \in R^{p'}} \sum_{s \in S \setminus F^t} b_{rs} \alpha_s^p > 0$, and thus $v_a(p', F^t) > v_a(p, F^t)$. This is impossible since $p \in \bar{P}_{\Delta^{B_a}}^{F^t}$. $\square$

Proposition 15. *If, for a given set F^t , p is an optimal extreme point of Δ^{B_a} such that there exists a short connection $s \in S \setminus F^t$ with $\alpha_s^p > 0$, then there exists another feasible point p' , where $\alpha_s^{p'} = \alpha_s^p$, for all $s \in F^t$, and $\alpha_s^{p'} = 0$, for all $s \in S \setminus F^t$, such that $v_a(p', F^t) = v_a(p, F^t)$.*

Proof. From Proposition 14, if, for a given set F^t , $p \in \bar{P}_{\Delta^{B_a}}^{F^t}$, then $\exists r_i \in R^p \mid \sum_{s \in F^t} b_{r_i s} \alpha_s^p = \beta^p$. Hence, even if $\alpha_s^{p'} \neq \alpha_s^p, \forall s \in S \setminus F^t, \beta^{p'} = \beta^p$. From that, and since $\alpha_s^{p'} = \alpha_s^p, \forall s \in F^t, v_a(p', F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p = v_a(p, F^t)$. $\square$

Let $p_1 = (\alpha^{p_1}, \beta^{p_1})$ be an extreme point of Δ^{B_a} . From Proposition 10, $\alpha_s^{p_1} = \bar{c}$ for at least one short connection $s \in S^{p_1}$. Let $p_2 = (\alpha^{p_2}, \beta^{p_2})$ be a point in Δ^{B_a} where (i) $\alpha_s^{p_2} = \max_{j \in S} \alpha_j^{p_1} = \bar{c}, \forall s \in (S^{p_1} \cap F^t)$, (ii) $\alpha_s^{p_2} = \alpha_s^{p_1}, \forall s \in (S^{p_1} \setminus F^t)$, and (iii) $\beta^{p_2} = \max_{r \in R} \sum_{s \in S^{p_2}} b_{rs} \alpha_s^{p_2}$. Let $S^p = S^{p_1} = S^{p_2}$ be the common set of positive variables. One can observe that p_2 is feasible since it satisfies all constraints (6.45)-(6.47) from BDSPa. For a given set F^t , recall that $v_a(p, F^t)$ is the value of the objective function (6.44) of BDSPa at the point p .

Proposition 16. *If, for a given set F^t , p_1 is an extreme point of Δ^{B_a} and p_2 is a different point which has the same set of positive variables as p_1 , but in which the values of all variables $\alpha_s, s \in (S^{p_2} \cap F^t)$, are equal to their common upper bound, $\bar{c}$, then, $v_a(p_2, F^t) \geq v_a(p_1, F^t)$. In addition, if no aircraft routing $r \in R^{p_1}$ contains all short connections $s_i \in F^t$ such that $\alpha_{s_i}^{p_1} < 1$, then $v_a(p_2, F^t) > v_a(p_1, F^t)$.*

Proof. If $\bar{c} = 1$, then $\alpha_s^{p_2} = 1, \forall s \in (S^p \cap F^t)$. Let $r_1 \in R^{p_1}$ be one active constraint at p_1 and $S_1^p = \{s \in S^p \mid b_{r_1 s} = 1\}$ be the set of short connections from S^p included

in r_1 ($|R^{p_1}| \geq 1$ from Proposition 10). Let r_2 and S_2^p correspond to p_2 as r_1 and S_1^p correspond to p_1 . With the active constraint r_1 , one can calculate $\beta^{p_1} = \sum_{s \in S_1^p} \alpha_s^{p_1}$ and obtain $v_a(p_1, F^t) = \sum_{s \in F^t} \alpha_s^{p_1} - \beta^{p_1} = \sum_{s \in F^t} \alpha_s^{p_1} - \sum_{s \in S_1^p} \alpha_s^{p_1}$. The same is true for $v_a(p_2, F^t)$. Although $S^{p_1} = S^{p_2}$, it is easy to see that R^{p_2} can differ from R^{p_1} since there exists at least one short connection $s_i \in (S^p \cap F^t)$ with $\alpha_{s_i}^{p_1} \neq \alpha_{s_i}^{p_2}$. Only two possible situations can occur: either there is an active constraint $r_i \in R^{p_1}$ that is also active at p_2 , or there is none.

Assume that $r_1 \in (R^{p_1} \cap R^{p_2})$. In this case, $\beta^{p_2} = \beta^{p_1} + \sum_{s \in (S_1^p \cap F^t)} (1 - \alpha_s^{p_1})$. Therefore, $v_a(p_2, F^t) = v_a(p_1, F^t) + \sum_{s \in F^t} (1 - \alpha_s^{p_1}) - \sum_{s \in (S_1^p \cap F^t)} (1 - \alpha_s^{p_1})$. Since (i) $\alpha_s^{p_1} \leq 1, \forall s \in S^{p_1}$, (ii) $\exists s_i \in F^t | \alpha_{s_i}^{p_1} < 1$ and (iii) $(S_1^p \cap F^t) \subset F^t$, then $v_a(p_2, F^t) \geq v_a(p_1, F^t)$. If, in addition, no aircraft routings $r \in R^{p_1}$ contain all short connections $s_i \in F^t$ such that $\alpha_{s_i}^{p_1} < 1$, then $\sum_{s \in F^t} (1 - \alpha_s^{p_1}) > \sum_{s \in (S_1^p \cap F^t)} (1 - \alpha_s^{p_1})$ and $v_a(p_2, F^t) > v_a(p_1, F^t)$.

Assume that $|R^{p_1} \cap R^{p_2}| = 0$. In that case, $\beta^{p_2} = \beta^{p_1} + \sum_{s \in (F^t \cap S_1^p \cap S_2^p)} (1 - \alpha_s^{p_1}) - \sum_{s \in (S_1^p \setminus S_2^p)} \alpha_s^{p_1} + \sum_{s \in (S_2^p \setminus S_1^p)} \alpha_s^{p_2}$. Therefore, $v_a(p_2, F^t) = v_a(p_1, F^t) + \sum_{s \in F^t} \alpha_s^{p_2} - \sum_{s \in F^t} \alpha_s^{p_1} - \sum_{s \in S_2^p} \alpha_s^{p_2} + \beta^{p_1}$. Since r_2 is not active at p_1 , $\sum_{s \in S_2^p} \alpha_s^{p_1} < \beta^{p_1}$. Hence, $\sum_{s \in F^t} \alpha_s^{p_1} - \sum_{s \in (F^t \setminus S_2^p)} \alpha_s^{p_1} + \sum_{s \in (S_2^p \setminus F^t)} \alpha_s^{p_1} < \beta^{p_1}$. Seeing that (i) $\sum_{s \in (F^t \setminus S_2^p)} \alpha_s^{p_2} > \sum_{s \in (F^t \setminus S_2^p)} \alpha_s^{p_1}$, (ii) $\sum_{s \in (F^t \setminus S_2^p)} \alpha_s^{p_2} = \sum_{s \in F^t} \alpha_s^{p_2} - \sum_{s \in (F^t \cap S_2^p)} \alpha_s^{p_2}$, (iii) $\sum_{s \in (S_2^p \setminus F^t)} \alpha_s^{p_1} = \sum_{s \in (S_2^p \setminus F^t)} \alpha_s^{p_2}$ and (iv) $\sum_{s \in (F^t \cap S_2^p)} \alpha_s^{p_2} + \sum_{s \in (S_2^p \setminus F^t)} \alpha_s^{p_2} = \sum_{s \in S_2^p} \alpha_s^{p_2}$, then $\sum_{s \in F^t} \alpha_s^{p_1} - \sum_{s \in F^t} \alpha_s^{p_2} + \sum_{s \in S_2^p} \alpha_s^{p_2} < \beta^{p_1}$ and $v_a(p_2, F^t) > v_a(p_1, F^t)$. $\square$

To illustrate Proposition 16, recall that ray $c = \{1, 1, 0, 0, 1, 2, 0, 2\}$, from Example 1, is an extreme ray of Δ_1 . Yet, c is not a feasible point of $\Delta_1^{B_a}$ since it violates one of the upper bound constraints (6.46), but the equivalent extreme ray $l = \{1/2, 1/2, 0, 0, 1/2, 1, 0, 1\}$ is an extreme point $\Delta_1^{B_a}$. However, one can see that l is not always an optimal solution of BDSPa. For example, when $F^t = \{A, B, E, F\}$,

the value of the objective function (6.44) at l is lower than it is at the point $e = \{1, 1, 0, 0, 1, 1, 0, 2\}$ ($v_a(l, F^t) = 3/2 < v_a(e, F^t) = 2$), but e gives a weaker cut than l . The point e is, in fact, a linear combination of the extreme rays l and $b = \{1, 1, 0, 0, 1, 0, 0, 2\}$.

Proposition 17. *If, for a given set F^t , p is an optimal extreme point of Δ^{B_a} , then $\alpha_s^p = \bar{c}$, for all $s \in S^p$, or there exists another feasible point q , where $\alpha_s^q = \bar{c}$, for all $s \in S^q$, such that $v_a(q, F^t) = v_a(p, F^t)$, i.e., in at least one of the optimal points of $BDSP_a$, all positive variables α take the value of their common upper bound, $\bar{c}$.*

Proof. From Proposition 16, if $p \in P_{\Delta^{B_a}}$ and, for a given set F^t , there exists a short connection $s_i \in (S^p \cap F^t)$ with $\alpha_{s_i}^p < \bar{c}$, then, there exists another feasible point p' where (i) $\alpha_s^{p'} = \bar{c}, \forall s \in (S^p \cap F^t)$, (ii) $\alpha_s^{p'} = \alpha_s^p, \forall s \in (S^p \setminus F^t)$, and (iii) $v_a(p', F^t) \geq v_a(p, F^t)$. From that, and since p is optimal, $v_a(p', F^t) = v_a(p, F^t)$. On the one hand, if there are no short connections $s \in S \setminus F^t$ with $\alpha_s^p > 0$, then $(S^p \cap F^t) = S^p$ (or $(S^{p'} \cap F^t) = S^{p'}$). On the other hand, if there exists a short connection $s \in S \setminus F^t$ with $\alpha_s^p > 0$, then, from Proposition 15, since p (or p') is an optimal point in Δ^{B_a} , there exists another optimal point p'' where $\alpha_s^{p''} = 0, \forall s \in S \setminus F^t$. Hence, if $p \in \bar{P}_{\Delta^{B_a}}^{F^t}$, then $\alpha_s^p = \bar{c}, \forall s \in S^p$, or there exists another feasible point q , where $\alpha_s^q = \bar{c}, \forall s \in S^q$, such that $v_a(q, F^t) = v_a(p, F^t)$. $\square$

Proposition 18. *For a given short connection set F^t , cuts (6.43) correspond to an optimal extreme point of Δ^{B_a} .*

Proof. Cuts (6.43) correspond to a point in Δ^{B_a} , where $\bar{\alpha}_s = 1, \forall s \in F^t$, and $\bar{\beta} = |S(\bar{r}^t) \cap F^t|$. When $\bar{c} = 1$, the point is feasible and there are exactly $|F^t|$ active constraints (6.46) at $(\bar{\alpha}, \bar{\beta})$. In addition, at least one constraint (6.45) is also active at that point since, from the definition of β , there exists at least one routing using exactly β short connections. Cuts (6.43) thus correspond to an extreme point of Δ^{B_a} .

since $n(H^p) \geq |S^p| + 1$ (see Proposition 8). The objective function (6.44) chooses a solution that maximizes the difference between the number of positive variables from F^t in the solution and the value of β . One cannot increase the number of positive variables from F^t in cuts (6.43) since they are all already included, and a decrease in the number of variables from F^t in the cut would result in a smaller or equal decrease in the value of β . Cuts (6.43) thus correspond to an optimal extreme point of Δ^{B_a} . $\square$

Proposition 19. *Cuts (6.43) are the weakest of the cuts corresponding to the optimal solutions of BDSPa, according to the ratio violation/right-hand-side (v/rhs).*

Proof. For a given set F^t , define E as the set of cuts corresponding to an optimal solution of BDSPa. Let $lhsf_e$ be the sum of the coefficients of the short connections $s \in F^t$ in feasibility cut $e \in E$, and rhs_e be the value of the right-hand-side of cut e . According to the ratio v/rhs , the strength of the cut e is $rv_e = (lhsf_e + add_e - rhs_e)/rhs_e$, where add_e is the sum of the coefficients of $s \in S \setminus F^t$ in e . Let the element $e_1 \in E$ coincide with cut (6.43). Since $add_{e_1} = 0$, $rv_{e_1} = (|F^t| - rhs_{e_1})/rhs_{e_1}$ (all coefficients are equal to 1 in e_1 and $lhsf_{e_1} = |F^t|$). Recall that $v_a^*(F^t) = lhsf_e - rhs_e, \forall e \in E$. Since all objective values for solutions corresponding to a cut $e \in E$ are equal, then, for a given $e \in E$, $v_a^*(F^t) = lhsf_e - rhs_e = |F^t| - rhs_{e_1}$. From that, $rhs_e = rhs_{e_1} - |F^t| + lhsf_e$ and $rv_e = (lhsf_e + add_e - (rhs_{e_1} - |F^t| + lhsf_e))/(rhs_{e_1} - |F^t| + lhsf_e) = (add_e + |F^t| - rhs_{e_1})/(rhs_{e_1} - (|F^t| - lhsf_e))$. Since (i) $add_e \geq 0$, (ii) $add_{e_1} = 0$, (iii) $|F^t| \geq lhsf_e, \forall e \in E$, and (iv) $|F^t| = lhsf_{e_1}$, then $rv_e \geq rv_{e_1}, \forall e \in E$. $\square$

To illustrate Proposition 19, consider Example 1. If $F^t = \{C, E, F\}$, then cut (6.43) is $j_1 : x_C + x_E + x_F \leq 2$, but the following cuts also correspond to optimal solutions of BDSPa in that case: $j_2 : x_C + x_E \leq 1$, $j_3 : x_E + x_F \leq 1$, $j_4 : x_C + x_E + x_F + x_G \leq 2$, $j_5 : x_C + x_E + x_G \leq 1$ and $j_6 : x_E + x_F + x_G \leq 1$. All these cuts correspond to a

solution to BDSPa with an objective value of $v_a^*(F^t) = 1$ but with different strengths according to the ratio v/rhs since $rv_1 = 1/2$, $rv_2 = 1$, $rv_3 = 1$, $rv_4 = 1$, $rv_5 = 2$ and $rv_6 = 2$. One can notice that j_2 , j_3 , j_5 and j_6 are extreme rays of Δ_1 and j_2 and j_3 are MIS feasibility cuts (6.29).

6.4.2 Model (b): A unique artificial variable appearing in every short connection linking constraint

When model (b) is chosen, only one artificial variable is added in the primal subproblem and the bounded dual subproblem (BDSPb) is the following:

$$\text{Maximize } \sum_{k \in K} \sum_{s \in S} d_{ks} \bar{x}_k \alpha_s - \beta \quad (6.52)$$

subject to

$$\sum_{s \in S} b_{rs} \alpha_s - \beta \leq 0 \quad (r \in R) \quad (6.53)$$

$$\sum_{s \in S} \alpha_s \leq 1 \quad (6.54)$$

$$\alpha_s \geq 0 \quad (s \in S). \quad (6.55)$$

No new dual variable is introduced, but the sum of the α_s variables is now bounded by the value 1. Let Δ^{B_b} denote the polyhedron defined by constraints (6.53)-(6.55) and let $P_{\Delta^{B_b}}$ be the set of extreme points of Δ^{B_b} .

One can easily see that Propositions 8, 9, 14 and 15, from model (a), can be generalized to model (b) since the set of dual variables, the set of dual constraints (6.45) and the

dual subproblem objective function remain the same. Further properties may be studied.

Proposition 20. *If p is an extreme point of Δ^{B_b} , then $\sum_{s \in S^p} \alpha_s^p = 1$ (constraint (6.54) is active), and there are exactly $|S^p|$ linearly independent active constraints (6.53) at p .*

Proof. If $p \in P_{\Delta^{B_b}}$, then $n(H^p) \geq |S^p| + 1$ in BDSPb (see Proposition 8). From Proposition 9, there cannot be more than $|S^p|$ linearly independent active constraints (6.53). From that, and since there is only one constraint (6.54), the latter constraint is active and there are exactly $|S^p|$ linearly independent active constraints (6.53) at p in BDSPb. $\square$

Proposition 21. *If q is an extreme ray of Δ , then the point p , with (i) $\alpha_s^p = \alpha_s^q / \sum_{s \in S^q} \alpha_s^q$, $\forall s \in S^q$, (ii) $\alpha_s^p = 0$, $\forall s \in S \setminus S^q$, and (iii) $\beta^p = \beta^q / \sum_{s \in S^q} \alpha_s^q$, is an extreme point of Δ^{B_b} , i.e., every extreme ray of Δ corresponds to an extreme point of Δ^{B_b} , which is a rescaling of the ray such that constraint (6.54) is satisfied as an equality.*

Proof. If $q \in R_\Delta$, then $n(R^q) = |S^q|$ in DSP. By definition, $\sum_{s \in S^q} b_{rs} \alpha_s^q = \beta^q$, $\forall r \in R^q$. Since $S^p = S^q$, $\sum_{s \in S^p} b_{rs} \alpha_s^p = \sum_{s \in S^q} b_{rs} (\alpha_s^q / \sum_{s \in S^q} \alpha_s^q) = \beta^q / \sum_{s \in S^q} \alpha_s^q = \beta^p$, $\forall r \in R^q$. Therefore, $R^p = R^q$ and $n(R^p) = |S^p|$ in BDSPb. In addition, constraint (6.54) is satisfied as an equality since $\sum_{s \in S} \alpha_s^p = \sum_{s \in S^q} (\alpha_s^q / \sum_{s \in S^q} \alpha_s^q) = 1$. Hence, $n(H^p) = |S^p| + 1$ in BDSPb and, from Proposition 9, $p \in P_{\Delta^{B_b}}$. $\square$

To illustrate Proposition 21, recall that the ray $b = \{1, 1, 0, 0, 1, 0, 0, 2\}$, from Example 1, is an extreme ray of Δ_1 and it corresponds to the feasibility cut $x_A + x_B + x_E \leq 2$. From the rescaling described in the proposition, one can construct point $b' = \{1/3, 1/3, 0, 0, 1/3, 0, 0, 2/3\}$. One can verify that the same set of dual

constraints (6.53) are active at b and b' , and that constraint (6.54) is satisfied as an equality at b' . Point b' corresponds to the cut $1/3 x_A + 1/3 x_B + 1/3 x_E \leq 2/3$, which is, in fact, exactly the same as the cut generated from b .

Proposition 22. *All extreme points of Δ^{B_b} correspond to an extreme ray of Δ , and conversely.*

Proof. Every extreme ray of Δ corresponds to an extreme point of Δ^{B_b} (see Proposition 21). Hence, $R_\Delta \subset P_{\Delta^{B_b}}$. From Proposition 20, if $p \in P_{\Delta^{B_b}}$, then $n(R^p) = |S^p|$ in BDSPb. Since the set of constraints (6.12) is the same as the set (6.53), $n(R^p) = |S^p|$ also in DSP and $P_{\Delta^{B_b}} \subset R_\Delta$. Hence, $P_{\Delta^{B_b}} = R_\Delta$. $\square$

Let $q_1 = (\alpha^{q_1}, \beta^{q_1})$ and $q_2 = (\alpha^{q_2}, \beta^{q_2})$ be two extreme rays of Δ such that q_2 yields a stronger cut than q_1 , i.e., $\beta^{q_2} = \beta^{q_1}$ and $\sum_{s \in S} \alpha_s^{q_2} > \sum_{s \in S} \alpha_s^{q_1}$. Let $p_1 = (\alpha^{p_1}, \beta^{p_1})$ and $p_2 = (\alpha^{p_2}, \beta^{p_2})$ be the two extreme points of Δ^{B_b} corresponding to q_1 and q_2 , respectively, i.e., rescalings of q_1 and q_2 such that $\sum_{s \in S} \alpha_s = 1$ (see Proposition 21). For a given set F^t , let $v_b(p, F^t)$ be the value of the objective function (6.52) of BDSPb at the point p .

Proposition 23. *If q_1 and q_2 are two extreme rays of Δ such that q_2 yields a stronger cut than q_1 , and if p_1 and p_2 are the two corresponding extreme points of Δ^{B_b} , then, for a given short connection set F^t , $v_b(p_2, F^t) > v_b(p_1, F^t)$, if, in addition, $\alpha_s^{q_2} = 0$, for all $s \in S \setminus F^t$.*

Proof. Recall that $v_b(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p$. From Proposition 21, $\beta^{p_1} = \beta^{q_1} / \sum_{s \in S^{q_1}} \alpha_s^{q_1}$ and $\beta^{p_2} = \beta^{q_2} / \sum_{s \in S^{q_2}} \alpha_s^{q_2}$. Since, by definition, $\beta^{q_1} = \beta^{q_2}$ and $\sum_{s \in S^{q_1}} \alpha_s^{q_1} < \sum_{s \in S^{q_2}} \alpha_s^{q_2}$, then $\beta^{p_2} < \beta^{p_1}$. From Proposition 20, $\sum_{s \in S} \alpha_s^{p_1} = \sum_{s \in S} \alpha_s^{p_2} = 1$. If $\alpha_s^{q_2} = 0, \forall s \in S \setminus F^t$, then $\sum_{s \in F^t} \alpha_s^{p_2} = 1 \geq \sum_{s \in F^t} \alpha_s^{p_1}$. Hence, $v_b(p_2, F^t) > v_b(p_1, F^t)$. $\square$

To illustrate Proposition 23, recall that the ray $b = \{1, 1, 0, 0, 1, 0, 0, 2\}$ and the ray $c = \{1, 1, 0, 0, 1, 2, 0, 2\}$, from Example 1, are two extreme rays of Δ_1 and they correspond to the feasibility cuts $x_A + x_B + x_E \leq 2$ and $x_A + x_B + x_E + 2x_F \leq 2$, respectively. One can observe that the cut generated from c is strictly stronger than the cut generated from b . From Proposition 21, one can construct the points $b' = \{1/3, 1/3, 0, 0, 1/3, 0, 0, 2/3\}$ and $c' = \{1/5, 1/5, 0, 0, 1/5, 2/5, 0, 2/5\}$ in $\Delta_1^{B_b}$ which correspond to b and c , respectively. If $F^t = \{A, B, E, F\}$, then one can easily see that the stronger cut would have been generated first since $v_b(c', F^t) = 3/5 > v_b(b', F^t) = 1/3$. However, if $F^t = \{A, B, E\}$, $v_b(c', F^t) = 1/5 < v_b(b', F^t) = 1/3$.

6.4.3 Model (c): A unique artificial variable appearing in the convexity constraint

To obtain BDSPc, only one constraint is added to DSP in order to bound it: $\beta \leq 1$. One can easily verify that to every extreme point of the polyhedron of BDSPb (resp. BDSPc) corresponds an extreme point of the polyhedron of BDSPc (resp. BDSPb), which is a rescaling of the point such that $\beta = 1$ (resp. $\sum_{s \in S} \alpha_s = 1$). The form of BDSPc is thus similar to the form of BDSPb. One can also verify that all results that hold for model (b) also hold for model (c). Model (c) will thus not be part of the discussion in the remainder of the paper.

6.4.4 Comparison of the different bounding approaches for the Benders dual subproblem

Propositions 11 and 21 indicate that the sets of Benders cuts generated from BDSPa and BDSPb both include all feasibility cuts (6.16) generated from extreme rays of DSP. Proposition 22 indicates that the set of cuts generated from BDSPb is, in fact, the

same as the set of feasibility cuts (6.16) generated from DSP. However, by Proposition 13, the set of cuts generated from BDSPa also includes cuts that do not correspond to extreme rays (they correspond to rays which are combinations of extreme rays). Although these extra cuts yield valid feasibility cuts, a larger number of iterations could be required before converging to an optimal solution of the integrated problem (6.1)-(6.6). In addition, by Propositions 17 and 23, for a given short connection set chosen by the crew pairing problem at iteration t , F^t , model (a) is likely to generate weaker cuts than model (b).

6.4.4.1 Cuts with a right-hand-side of value zero

Recall that S was defined, in Section 6.1, as the set of short connections included in at least one aircraft routing solution. When the ECP formulation (6.1)-(6.6) is chosen and complete aircraft routing solutions are included in the formulation, removing from the set of short connections the ones that are individually infeasible for the aircraft routing problem is an easy task. Indeed, all the arcs corresponding to a short connection s such that $b_{r,s} = 0$, for all $r \in R$, can be removed from the crew networks in a preprocessing step. However, getting all feasible routings is, by itself, a very difficult combinatorial optimization problem (MERCIER *et al.* (2005)). When complete routing solutions are not available and an explicit formulation is used for the integrated problem, determining the individually infeasible short connection set requires solving a series of aircraft routing problems, imposing the use of every short connection, one at a time. If this preprocessing step is not done (it can be very time-consuming), arcs corresponding to individually infeasible short connections for the aircraft cannot be removed from the crew networks and the set of short connections S includes a subset of individually infeasible short connections, *i.e.*, short connections which are not included in any feasible aircraft routing. In that case, some feasibility

cuts forbidding their use in the crew pairings will be generated. These cuts have a right-hand-side of value zero and they correspond to a dual solution in which $\beta = 0$. Interestingly, model (a) has the ability to generate cuts forbidding, at the same time, the use of all individually infeasible short connections taken by the crews at a given iteration, whereas model (b) does not. Therefore, although model (b) directly finds extreme rays of the original Benders subproblem (DSP), in practice, model (a) may dominate model (b).

Let $\hat{S}$ be the set of short connections not included in any aircraft routing solution. When $|S \cap \hat{S}| \neq 0$, i.e., the set of short connections S includes a subset of infeasible short connections, constraints (6.45) and (6.53) from BDSPa and BDSPb are in fact:

$$\sum_{s \in S \setminus \hat{S}} b_{rs} \alpha_s - \beta \leq 0 \quad (r \in R). \quad (6.56)$$

Recall that, for a given set F^t , (i) $v_a(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p$ is the value of the objective function (6.44) of BDSPa at the point p , (ii) $v_a^*(F^t)$ is the optimal value of BDSPa, and (iii) $\bar{P}_{\Delta^{B_a}}^{F^t} = \{p \in P_{\Delta^{B_a}} \mid v_a(p, F^t) = v_a^*(F^t)\}$ is the set of optimal extreme points of Δ^{B_a} . Also recall that we can assume, w.l.o.g., that the cost of using an artificial variable in an aircraft routing solution is 1, i.e., $\bar{c} = 1$.

Proposition 24. *For a given set of short connections F^t , if p is an optimal point of Δ^{B_a} where $\beta^p = 0$, then $\alpha_s^p = 1, \forall s \in (\hat{S} \cap F^t)$ and p is an extreme point of Δ^{B_a} , i.e., if there is an optimal solution to BDSPa in which $\beta = 0$, then all variables from F^t corresponding to an infeasible short connection take the value 1 in the solution and it corresponds to an extreme point.*

Proof. When $\beta^p = 0$, then $\alpha_s^p = 0, \forall s \in (S \setminus \hat{S})$, and $v_a(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p = \sum_{s \in F^t} \alpha_s^p = \sum_{s \in (\hat{S} \cap F^t)} \alpha_s^p$. Hence, increasing the value of the variables corresponding to the short connections $s \in (\hat{S} \cap F^t)$ directly improves the value of the objective

function, and this, without violating any dual constraints (6.56). From that, and since all variables $\alpha_s^p, s \in \hat{S}$, are individually bounded by $\bar{c} = 1$ (constraints (6.46)), then, $\alpha_s^p = 1, \forall s \in (\hat{S} \cap F^t)$, when $\beta^p = 0$ and $v_a(p, F^t) = v_a^*(F^t)$. In addition, since the total number of linearly independent active constraints at p ($|S^p|$ constraints (6.46) as well as $|S| - |S^p| + 1$ non-negativity constraints) is equal to the total number of variables ($|S| + 1$), then p is an extreme point of Δ^{B_a} . $\square$

Proposition 25. *For a given set of short connections F^t , if $|\hat{S} \cap F^t| \neq 0$ and p is an optimal extreme point of Δ^{B_b} where $\beta^p = 0$, then $|S^p| = 1$, i.e., only one variable from F^t corresponding to an infeasible short connection takes a positive value.*

Proof. When $\beta^p = 0$, then $\alpha_s^p = 0, \forall s \in (S \setminus \hat{S})$, and $v_b(p, F^t) = \sum_{s \in F^t} \alpha_s^p - \beta^p = \sum_{s \in F^t} \alpha_s^p = \sum_{s \in (\hat{S} \cap F^t)} \alpha_s^p$. Hence, increasing the value of the variables corresponding to the short connections $s \in (\hat{S} \cap F^t)$ directly increases the value of the objective function without violating constraints (6.56), but, by constraint (6.54), $\sum_{s \in S} \alpha_s^p \leq 1$. From that, and since $v_b(p, F^t) = v_b^*(F^t)$, then $\sum_{s \in (\hat{S} \cap F^t)} \alpha_s^p = 1$ and $v_b(p, F^t) = \sum_{s \in (\hat{S} \cap F^t)} \alpha_s^p = 1$. One can observe that p is an extreme point of Δ^{B_b} only if $|S^p| = 1$, since the total number of active constraints (constraint (6.54) as well as $|S| - |S^p| + 1$ non-negativity constraints) is equal to the total number of variables ($|S| + 1$) only in that case. $\square$

Proposition 24 implies that, for a given crew pairing solution, a single cut simultaneously forbidding all maintenance infeasible short connections from the set chosen by the crew pairings, F^t , is generated by model (a). In contrast, Proposition 25 indicates that as many as $|\hat{S} \cap F^t|$ cuts are generated by model (b) to forbid every short connection from the same set. One can easily see that the cuts with a zero right-hand-side generated by model (b), i.e. cuts with only one short connection, correspond to extreme rays of Δ in the direction of the non-negativity constraints of the positive short connections. Similarly, the cut generated by model (a) corresponds to a linear

combination of all extreme rays of Δ in the direction of the non-negativity constraints of the positive short connections. Hence, in the case of cuts with a right-hand-side of value zero, linear combinations of extreme rays give stronger cuts than extreme rays. In fact, the numerical results reported in Section 6.6 show that the number of feasibility cuts generated with model (b) is smaller than with model (a) only when the individually infeasible short connections arcs are removed from the model. In addition, when an explicit formulation is used for the integrated problem, the total time needed to solve model (a) is even smaller than the time needed to identify the individually infeasible short connections. Recall that, from Proposition 3, the MIS feasibility cuts (6.29) correspond to extreme rays of Δ . These cuts could therefore not generate the stronger zero right-hand-side cuts when the infeasible arcs are kept in the crew networks.

6.5 Strengthening the Feasibility Cuts

By definition, at iteration t , feasibility cuts (6.29) and (6.43) never include short connections which are not part of the set used by the optimal crew pairing solution at iteration t , F^t . By Proposition 15, even though it is possible to include some short connections in $S \setminus F^t$ when generating a cut by solving BDSPa or BDSPb, neither problem favors them since these short connections do not contribute to the objective function of the problem. This section proposes two ways of strengthening feasibility cuts. The first procedure can be applied to Benders cuts generated from BDSPa or BDSPb, whereas the second one can be applied to any feasibility cut for the integrated aircraft routing and crew pairing problem.

6.5.1 Pareto-optimal cuts

Whenever the primal subproblem (6.38)-(6.41) is degenerate (and this is often the case in practice), there may exist more than one optimal solution to the bounded dual subproblem. Although any of these points leads to a valid optimality cut, some may yield stronger cuts than others. The cut generated from the extreme point (α^1, β^1) dominates the cut generated from the extreme point (α^2, β^2) if and only if

$$\sum_{k \in K} \sum_{s \in S} d_{ks} \alpha_s^1 x_k - \beta^1 \geq \sum_{k \in K} \sum_{s \in S} d_{ks} \alpha_s^2 x_k - \beta^2$$

for all $\bar{x} \in X$ with strict inequality for at least one point. A cut is said to be Pareto-optimal if no other cuts dominate it (MAGNANTI et WONG, 1981).

Let X^{LP} be the polyhedron defined by (6.2) and constraints $x_k \geq 0$ ($k \in K$), and let $ri(X^{LP})$ denote the relative interior of X^{LP} . Pareto-optimal cuts are generated from extreme points, one thus has to ensure that the dual subproblem is bounded for any solution of the relaxed master problem. For a given solution $\bar{x} \in X^{LP}$, let $v(\bar{x})$ denote the optimal value of the primal subproblem. To identify an optimal solution to the bounded dual subproblem that yields a Pareto-optimal cut, one must solve the following auxiliary bounded dual subproblem (ABDSP), where x^0 is a point chosen in $ri(X^{LP})$:

$$\text{Maximize } \sum_{k \in K} \sum_{s \in S} d_{ks} x_k^0 \alpha_s - \beta \quad (6.57)$$

subject to

$$\sum_{k \in K} \sum_{s \in S} d_{ks} \bar{x}_k \alpha_s - \beta = v(\bar{x}) \quad (6.58)$$

$$(\alpha, \beta) \in \Delta^B. \quad (6.59)$$

Besides including the same bounded dual subproblem constraints (6.59) ($\Delta^B = \Delta^{B_a}$ or Δ^{B_b}), the ABDSP contains an additional constraint (6.58) to ensure that one will choose an extreme point from the set of optimal solutions to the original bounded dual subproblem (with a value equal to the original optimal value $v(\bar{x})$). The auxiliary problem is solved at every iteration of the Benders decomposition algorithm, after BDSP has been solved. The objective function of the auxiliary problem (6.57) compares all possible cuts at a point x^0 in the feasible region of the master problem. It therefore tries to maximize the strength of the cut to be added to the Benders master problem.

When $\Delta^B = \Delta^{B_a}$ in (6.59), the weakness highlighted by Proposition 17 for cuts generated directly from BDSPa is also true for cuts generated from ABDSP because the upper bound constraints on the dual variables are still present in ABDSP. However, since x^0 is a point chosen in $ri(X^{LP})$, all dual variables α may have a positive coefficient in the objective function (6.57), even those corresponding to short connections $s \in S \setminus F^t$. Finding the Pareto-optimal cuts can thus reduce the number of cuts needed before the Benders decomposition algorithm converges.

6.5.2 Lifting procedure

The preceding section showed that the generation of Pareto-optimal Benders cuts may favor the inclusion of short connections which are not part of the set used by the optimal crew pairing solution at iteration t , F^t , but the strongest cuts may still be left out by solving ABDSP. Nevertheless, one can use a lifting procedure to strengthen all types of feasibility cuts. Cuts (6.29) and (6.43), as well as those generated from extreme rays of Δ or from extreme points of Δ^{B_a} or Δ^{B_b} , all correspond to feasible solutions of DSP since they all satisfy the dual constraints (6.12) and (6.13). The lifting procedure presented here consists of first generating a feasibility

cut with a chosen method and then lifting it by solving a modified version of DSP. Let $u = (\bar{\alpha}^u, \bar{\beta}^u)$ be the solution corresponding to a cut that one wishes to strengthen. From u , a modified Benders dual subproblem can be formulated as follows (MDSP(u)):

$$\text{Maximize } \sum_{s \in S} \alpha_s \quad (6.60)$$

subject to

$$\sum_{s \in S} b_{rs} \alpha_s \leq \bar{\beta}^u - \sum_{s \in S} b_{rs} \bar{\alpha}_s^u \quad (r \in R) \quad (6.61)$$

$$\alpha_s \geq 0 \quad (s \in S). \quad (6.62)$$

Constraints (6.61) limit the lifting of the dual variables α_s to the slack of the constraints (6.12) of DSP at u . The right-hand-sides of constraints (6.61) are thus all positive and bounded. This makes MDSP(u) a bounded problem (without the use of specific upper bounds on the variables) and also a feasible problem since the null vector, representing no strengthening, is always feasible. One can observe that the dual variable β is not included in MDSP(u). The objective function (6.60) maximizes the sum of all variables $\alpha_s, s \in S$. The coefficient of a variable in the lifted cut is the sum of its value in u and its value in an optimal solution of MDSP(u). Let Δ^L denote the polyhedron defined by constraints (6.61)-(6.62) and let P_{Δ^L} be the set of extreme points of Δ^L . Let $v = (\bar{\alpha}^v)$ be an optimal solution to MDSP(u), and $w = (\bar{\alpha}^u + \bar{\alpha}^v, \bar{\beta}^u)$ be the constructed point from which the lifted cut is generated. The cut corresponding to w is always stronger than or equal to the cut corresponding to u since the value of both right-hand-sides is equal and the coefficient of every short connection $s \in S$ in the left-hand-side of the lifted cut is larger than or equal to its corresponding value in the cut generated from u ($\bar{\alpha}_s^v \geq 0, \forall s \in S$).

One can observe that since all variables have a positive coefficient in the objective function (6.60), MDSP(u) favors the stronger cuts which include short connections $s \in S \setminus F^t$. In addition, the problem highlighted by Proposition 17 is not present in MDSP(u) because the upper bound constraints are removed. Lifted cuts can thus contain variables with different coefficients. To illustrate this last statement, recall that the ray $b = \{1, 1, 0, 0, 1, 0, 0, 2\}$, from Example 1, is an extreme ray of Δ_1 and it corresponds to the cut $x_A + x_B + x_E \leq 2$. When the lifting procedure is applied to b , *i.e.*, when $u = b$, MDSP(u) is the following:

$$\text{Maximize } \alpha_A + \alpha_B + \alpha_C + \alpha_D + \alpha_E + \alpha_F + \alpha_G \quad (6.63)$$

subject to

$$\alpha_A + \alpha_B + \alpha_C \leq 0 \quad (6.64)$$

$$\alpha_A + \alpha_B + \alpha_D \leq 0 \quad (6.65)$$

$$\alpha_A + \alpha_C + \alpha_D \leq 1 \quad (6.66)$$

$$\alpha_B + \alpha_C + \alpha_D \leq 1 \quad (6.67)$$

$$\alpha_B + \alpha_D + \alpha_E \leq 0 \quad (6.68)$$

$$\alpha_A + \alpha_E \leq 0 \quad (6.69)$$

$$\alpha_C + \alpha_F \leq 2 \quad (6.70)$$

$$+ \alpha_G \leq 2 \quad (6.71)$$

$$\alpha_A, \alpha_B, \alpha_C, \alpha_D, \alpha_E, \alpha_F, \alpha_G \geq 0. \quad (6.72)$$

One can observe that the problem can be simplified since variables $\alpha_A, \alpha_B, \alpha_C, \alpha_D, \alpha_E$ can be removed, as well as constraints (6.64)-(6.69). The optimal solution to the modified problem is $v = \{0, 0, 0, 0, 0, 2, 2, 0\}$, *i.e.*, $\alpha_F = 2$ and $\alpha_G = 2$. One can construct the ray $w = (\bar{\alpha}^u + \bar{\alpha}^v, \bar{\beta}^u) = \{1, 1, 0, 0, 1, 2, 2, 2\}$. The corresponding lifted

cut, $x_A + x_B + x_E + 2x_F + 2x_G \leq 2$, is stronger than the original one, and, since $n(R^w) = 5 = S^w$, w is an extreme ray of Δ_1 .

Proposition 26. *When the lifting procedure is applied to a cut corresponding to an extreme ray of Δ , the resulting cut also corresponds to an extreme ray of Δ .*

Proof. If u is an extreme ray of Δ , then, from Proposition 1, $n(R^u) = |S^u|$. From that, and since there is no slack in an active constraint, it is easy to see that $|S^u \cap S^v| = 0$, i.e., the variables that are positive in u take the value 0 in v . Hence, $\alpha_s^w = \alpha_s^u, \forall s \in S^u$, and $|S^w| = |S^u| + |S^v|$. Since v is an extreme point of Δ^L and there are no upper bound constraints and no variable β in the model, $n(R^v) = |S^v|$. Recall that $\sum_{s \in S} b_{rs} \bar{\alpha}_s^w = \sum_{s \in S} b_{rs} \bar{\alpha}_s^v + \sum_{s \in S} b_{rs} \bar{\alpha}_s^u$ and $\bar{\beta}^w = \bar{\beta}^u$. Then, for a given routing $r \in R$, $\sum_{s \in S} b_{rs} \bar{\alpha}_s^w - \bar{\beta}^w = 0$ if and only if $\sum_{s \in S} b_{rs} \bar{\alpha}_s^v = \bar{\beta}^u - \sum_{s \in S} b_{rs} \bar{\alpha}_s^u$. That is to say, a constraint (6.12) in DSP is active at w if and only if the corresponding constraint (6.61) in MDSP(u) is active at v . In addition, since $\alpha_s^v \geq 0$, all constraints in DSP that are active at u are also active at w , thus, $n(R^w) = n(R^u) + n(R^v)$. From this last statement and since (i) $|S^w| = |S^u| + |S^v|$, (ii) $|S^u| = n(R^u)$, (iii) $|S^v| = n(R^v)$, then $|S^w| = n(R^u) + n(R^v) = n(R^w)$ and the point w corresponds to an extreme ray of Δ . $\square$

Proposition 26 implies that any MIS cut and any cut generated by solving DSP or BDSPb can be lifted and the resulting cut necessarily corresponds to a stronger extreme ray of Δ , if one exists, with $\alpha_s = \alpha_s^u, \forall s \in S^u$. However, if the procedure is applied to a cut which is not an extreme ray of Δ , the resulting cut corresponds to a stronger ray of Δ , if one exists, but not necessarily to an extreme ray of Δ .

6.6 Computational Experiments

In this section, we present computational experiments that were performed on instances based on data provided by two major airlines. We first provide a description of these instances, followed by a summary of our computational experiments. These experiments were done to compare the different dual subproblem bounding strategies in a Benders decomposition method, and to evaluate the two proposed strengthening methods. We did not compare all families of feasibility cuts. In fact, from Proposition 19, it is clear that the Benders cuts dominate cuts (6.43). Propositions 3 and 4 show that all MIS cuts (6.29) are included in the Benders cuts and that some of the stronger Benders cuts do not correspond to MIS cuts. Furthermore, the generation of an MIS cut implies the solution of an optimization problem on top of the regular aircraft maintenance routing problem. We think that although this family of cuts can be interesting since it leaves out the weakest of the feasibility cuts, the generation procedure is too demanding and would offset the benefits. As a matter of fact, the authors themselves did not implement the proposed cuts (see COHN and BARNHART (2003)).

6.6.1 Description of data sets

The test instances come from daily fleet assignment solutions provided by the airlines. The characteristics of the different instances are summarized in Table 6.1. This table indicates the number of daily legs and the number of short connections (SC).

The maintenance aircraft routing problem is solved to determine the sequence of flight legs to be flown by each individual aircraft so as to cover each leg exactly once while ensuring appropriate aircraft maintenance with the available number of aircraft.

Table 6.1 Characteristics of test instances

	B757C	A320D	D9SA	D9SB	B767S	MD80L
<i>Legs</i>	184	258	523	508	510	707
<i>SC</i>	114	183	502	659	370	1183

Recall that the ECP formulation includes all *unique and maximal maintenance-feasible short connection sets* (UM), *i.e.*, columns containing all the information needed from the complete solutions to the maintenance aircraft routing problem. Prior to solving ECP, one thus has to solve a series of aircraft routing problems with additional constraints and a modified objective function to generate all UM columns.

The crew pairing problem is solved to determine a minimum-cost set of pairings so that every flight leg is assigned a qualified crew and each pairing follows a set of applicable work rules related to a large number of factors such as flight time, rest time, connection time, etc. One may observe that, to compare the feasibility cuts studied in this paper, the crew pairing part of the integrated problem can be modeled in any way. It is, in fact, independent of the feasibility cut procedure. For the experiments reported in this section, the feasibility constraints are modeled through the use of resources and are handled directly by dynamic programming within a column generation framework (see, *e.g.*, DESAULNIERS *et al.* (1998b)). Details on the crew cost function and on the work rules considered can be found in MERCIER *et al.* (2005).

6.6.2 Summary of computational experiments

The integrated aircraft routing and crew pairing problem has been solved using alternately BDSPa and BDSPb as the Benders dual subproblem, with either the basic solution process (*basic*), an auxiliary Pareto-optimal subproblem (*P-O*) or the proposed lifting procedure (*Li*). The different combinations have also been used jointly

with two refinements: removing, in a preprocessing step, all arcs in the crew networks corresponding to individually infeasible short connections in order to eliminate the need for the zero right-hand-side cuts (method *RA*), and removing, at every iteration, the linking constraints corresponding to the short connections not chosen by the current crew pairing solution in order to make the subproblem easier to solve (method *RC*). Method *RC* forces the dual variables α corresponding to the removed constraints to take the value zero in an optimal solution. One can easily observe that the Benders cuts generated would likely be weaker, but still valid. Our algorithms were coded in C++ and all experiments were performed on a Sun UltraSPARC-II computer with a 480MHz processor.

Since the branch-and-bound methodology used for the integer crew pairing problem is heuristic (due to the size and difficulty of the problem) and the purpose of this section is to compare the different cut generation procedures, only the LP relaxation of the different instances have been solved. Tables 6.2 and 6.3 present a comparison of the number of Benders cuts and the CPU time needed to solve the LP relaxation when the arcs in the crew networks corresponding to infeasible short connections are either kept or removed in a preprocessing step (*RA*). One can observe that when the infeasible short connections are kept in the model and the lifting procedure is used, $\text{MDSP}(u)$ can be unbounded since some dual variables included in the objective function may not appear in any constraint. The results using the lifting procedure are thus only given when it is used jointly with method *RA*.

One can clearly see from Table 6.2 that when the maintenance infeasible short connection arcs are kept in the crew networks, the number of feasibility cuts needed is always smaller when *BDSPa* is used. Not only the number of cuts generated is smaller, but the first ones are also stronger. The latter statement is illustrated in Figure 6.1 from Appendix A, where the crew costs are plotted as a function of the number of iterations. One can see from the figure that the lower bounds given by

model (a) are always higher than the ones given by model (b). Instances 320D and MD80L were chosen as an illustration, but the same is true for all instances. As Propositions 24 and 25 showed, when a cut corresponding to a set of individually infeasible short connections (zero right-hand-side) was generated at a given iteration, all infeasible short connections in the set taken by the current crew solution indeed had a positive coefficient in the cut when model (a) was used whereas only one of them was included with model (b).

Table 6.2 Crew arcs corresponding to infeasible short connections are kept[†]

	B757C	A320D	D9SA	D9SB	B767S	MD80L
<i>Model (a)</i>						
<i>Cuts - Basic</i>	9	20	7	15	12	20
<i>CPU</i>	0.34	1.32	4.83	7.18	5.63	46.28
<i>Cuts - P-O</i>	7	19	2	6	10	14
<i>CPU</i>	0.38	1.61	4.17	5.21	5.40	28.42
<i>Cuts - Basic RC</i>	13	28	19	24	14	51
<i>CPU</i>	0.47	2.00	10.08	10.25	6.64	87.23
<i>Cuts - P-O RC</i>	16	48	10	10	16	66
<i>CPU</i>	0.66	3.25	6.88	6.92	7.88	98.98
<i>Model (b)</i>						
<i>Cuts - Basic</i>	23	52	101	146	56	216
<i>CPU</i>	0.67	3.04	56.03	83.14	20.86	376.51
<i>Cuts - P-O</i>	24	50	95	124	56	196
<i>CPU</i>	0.86	4.13	66.58	62.02	22.81	373.70
<i>Cuts - Basic RC</i>	24	49	93	139	61	198
<i>CPU</i>	0.75	3.52	51.30	59.93	20.53	395.31
<i>Cuts - P-O RC</i>	26	44	91	135	56	210
<i>CPU</i>	1.05	3.83	63.43	74.59	21.40	394.40

[†] All CPU times are in minutes.

When BDSPa is used, Table 6.2 shows that the number of cuts needed is greater when RC is used, but it is not always true when BDSPb is used. However, Figure 6.1 from Appendix A shows that even in the cases where the number of iterations needed is not greater with RC, the generated cuts are generally weaker. In fact, the dual variables corresponding to the removed constraints are not always null. Keeping them thus strengthens the cuts and, except for instance D9SB, the CPU times found in the table show that it is not compensated by the solution time saved by a reduced

number of constraints in the Benders primal subproblem. One can also see from the table that generating Pareto-optimal cuts (P-O) helps in reducing the number of iterations needed.

The comparison of Table 6.3 with Table 6.2 confirms that the number of iterations needed is reduced with RA. In addition, Table 6.3 and Figure 6.2 from Appendix A show that, with method RA, model (b) is generally better than model (a). When no zero right-hand-side cut is generated, extreme rays indeed yield stronger cuts than combinations of extreme rays. Since method RC did not yield better results when jointly used with method RA, the results for RC are left out of Table 6.3.

Table 6.3 Crew arcs corresponding to infeasible short connections are removed (RA)[†]

	B757C	A320D	D9SA	D9SB	B767S	MD80L
<i>Model (a)</i>						
<i>Cuts - Basic</i>	7	17	2	29	7	11
<i>CPU</i>	0.31	1.02	3.41	8.01	4.20	17.30
<i>Cuts - P-O</i>	5	15	2	9	8	16
<i>CPU</i>	0.34	1.23	3.32	4.36	4.55	26.97
<i>Cuts - Li</i>	6	15	2	20	7	9
<i>CPU</i>	0.31	1.02	3.24	7.01	3.87	17.19
<i>Model (b)</i>						
<i>Cuts - Basic</i>	12	13	2	10	7	13
<i>CPU</i>	0.39	0.94	3.27	5.33	3.74	20.60
<i>Cuts - P-O</i>	15	10	2	8	10	14
<i>CPU</i>	0.59	0.93	3.29	4.17	4.92	23.27
<i>Cuts - Li</i>	7	14	2	10	5	8
<i>CPU</i>	0.33	0.97	3.16	4.96	3.49	13.06

[†] All CPU times are in minutes.

When method RA is used, Table 6.3 shows that the Pareto-optimal cuts, as well as the lifted cuts (Li), help in reducing the number of iterations needed. In fact, the Pareto-optimal cuts are generally outperformed by the proposed lifting procedure in terms of speed of generation, and, in the case of model (b), also in terms of the number of cuts generated. In addition, since the lifting procedure takes less time than solving the Pareto-optimal auxiliary problem, one can see in Figure 6.3 from Appendix B

that even when the number of iterations with Li is larger than with P-O (instances A320B and D9SB), the lower bound given by the crew costs increases as fast with Li.

As Propositions 15 and 17 suggested, our experiments confirm that the cuts from model (a) show coefficients with equal values, and the cuts from both model (a) and model (b) do not include all incompatible short connections when they are not part of the set taken by the current crew solution. Our computational experiments also confirm that the proposed lifting procedure has the ability to overcome both weaknesses. The following example shows the first cut generated with model (a) for the fleet D9SA, and the corresponding lifted cut:

$$\text{Regular cut : } X_1 + X_2 + X_3 + X_4 \leq 3$$

$$\text{Lifted cut : } X_1 + X_2 + X_3 + X_4 + 2X_5 + X_6 \leq 3.$$

One can notice that a variable has a coefficient larger than one in the lifted cut. In our computational experiments, when they had a right-hand-side larger than zero, the lifted cuts and the cuts generated from model (b) often showed coefficients with different values. Using a short connection can indeed prevent the use of more than one other short connection (even up to four, in our experiments).

When the ECP formulation is used, and UM columns are included in the model, finding the set of individually infeasible short connections is an easy task: one only has to determine which short connections are not included in any column. In that case, model (b) is preferred to model (a) since the feasibility cuts generated from extreme rays of the dual subproblem polyhedron are stronger, and the lifting procedure is useful to strengthen the cuts and reduce the total computing time.

It is worth noting that some tests were done with the integrality constraints. A heuristic branch-and-bound method not only influences the computing time, but also

has an influence on the number of iterations of the Benders decomposition method. Nevertheless, we believe that the conclusions reached in our experiments can be extended to the integer formulation. The results for instance 757C are shown as an example in Table 6.4. One will notice that the number of cuts generated is higher with model (b) when RA is used. Although this seems to go against the conclusions drawn in the preceding sections, one can observe that it was also the case for this particular instance (and only this one) when the LP relaxation was solved (see Table 6.3).

Table 6.4 Instance 757C - Results for the integer formulation[†]

	<i>Basic (a)</i>	<i>P-O (a)</i>	<i>Li (a)</i>	<i>Basic (b)</i>	<i>P-O (b)</i>	<i>Li (b)</i>
<i>All arcs</i>						
<i>Cuts</i>	9	7	-	26	27	-
<i>CPU</i>	4.39	4.32	-	6.72	5.01	-
<i>Cost IP</i>	50 498	50 417	-	50 377	50 397	-
<i>RA</i>						
<i>Cuts</i>	7	6	5	13	15	7
<i>CPU</i>	0.59	1.49	2.99	2.32	0.99	0.53
<i>Cost IP</i>	50 425	50 275	50 425	50 418	50 353	50 417

[†] All CPU times are in minutes.

An integrated model with an explicit formulation

Finding all UM columns (or even a subset) is a difficult task (see MERCIER *et al.* (2005)), and for the larger problems, it can be more efficient to solve an integrated model which includes the maintenance aircraft routing formulation (*IMARF*) instead of complete aircraft routing solutions. Although the *IMARF* formulation is harder to solve, the increased solution time is well compensated by the time saved by not having to generate UM columns as in the *ECP* formulation. This section will show that the results of the comparison performed on the *ECP* formulation still hold for the *IMARF* formulation.

Although the aircraft maintenance constraints are modeled through the use of resources and are handled directly by dynamic programming in a column generation framework, the IMARF formulation includes additional constraints to ensure that all legs are covered by an aircraft and to limit number of available aircrafts. The Benders aircraft dual subproblem of the IMARF formulation thus contains more variables compared to the ECP formulation. The proposed lifting procedure is still valid in this case, but its efficiency is reduced since it only lifts the dual variables associated with the short connection linking constraints. We therefore did not experiment the lifting procedure with the IMARF formulation.

Table 6.5 presents a comparison of the number of Benders cuts and the CPU time needed to solve the LP relaxation of IMARF when the infeasible arcs are kept in the crew networks, while Table 6.6 gives the time needed to find all individually infeasible short connections (*preprocessing*). As it was the case with ECP, method RC did not prove to be useful and we therefore did not include the corresponding results in the tables.

Table 6.5 When the infeasible short connection arcs are kept - IMARF[†]

	B757C	A320D	D9SA	D9SB	B767S	MD80L
<i>Model (a)</i>						
<i>Cuts - Basic</i>	22	18	18	50	14	46
<i>CPU</i>	1.21	2.17	16.17	76.66	17.83	215.27
<i>Cuts - P-O</i>	15	12	14	21	11	1
<i>CPU</i>	1.92	3.40	21.98	100.16	42.60	68.33
<i>Model (b)</i>						
<i>Cuts - Basic</i>	32	45	85	61	36	85
<i>CPU</i>	1.62	4.87	53.15	47.96	29.24	241.10
<i>Cuts - P-O</i>	25	38	72	61	33	82
<i>CPU</i>	2.58	5.92	69.96	80.25	61.85	390.92

[†] All CPU times are in minutes.

Although it is not as blatant as with the ECP formulation, one can see from Table 6.5 that model (a) again dominates model (b) when the infeasible short connection arcs

are kept in the networks. The table also shows that the number of Pareto-optimal cuts ($P-O$) generated is smaller than the number of regular Benders cuts. Table 6.6 shows that the preprocessing step needed for method RA is more time consuming than the whole solution process when the arcs are kept. Hence, when the IMARF formulation is used, finding the set of individually infeasible short connections is a hard task and it is better to keep all arcs in the networks. In that case, model (a) is preferred to model (b).

Table 6.6 Identifying the individually infeasible short connections - IMARF[†]

	B757C	A320D	D9SA	D9SB	B767S	MD80L
<i>Preprocessing</i>	1.90	4.91	76.91	179.53	130.67	1612.80

[†] All CPU times are in minutes.

One can observe that, with this formulation, it is likely that a procedure using the MIS feasibility cuts (6.29) would spend much time generating all individually infeasible short connections one at a time since these cuts correspond to extreme rays of the Benders dual subproblem polyhedron. Finally, one can notice that the total computing time is not necessarily reduced when Pareto-optimal cuts are generated (solving the Pareto-optimal auxiliary problem is time-consuming), but systematically reducing the number of cuts might be preferred when the integer problem is solved.

6.7 Conclusion

When the integrated aircraft routing and crew pairing problem is solved either with Benders decomposition or another cut generation procedure, one has to verify at each iteration, in a subproblem, if the set of short connections used by the current crew solution is feasible for the aircraft routing problem. This paper has shown that there is no simple form for the extreme rays of the dual subproblem polyhedron. The dual

variables can indeed take different values in an extreme ray. In addition, feasibility cuts generated from predefined families (forcing each coefficient to take the value one) are all included in the set of Benders feasibility cuts, and may also be dominated by the latter.

To avoid having to identify extreme rays, one can make the Benders dual subproblem bounded. To achieve this, artificial variables must be introduced in the aircraft routing primal subproblem. This paper has compared different bounding methods and pointed out that the choice of artificial variables has an impact on the dual subproblem polyhedron, and thus, on the feasibility cuts generated. It was shown that cuts generated when a unique artificial variable is used correspond to extreme rays while it is not the case for cuts generated when a different artificial variable is added in each short connection linking constraint. The latter strategy may indeed favor cuts with equal coefficients even when they are weaker. In addition, all cuts do not necessarily include short connections that are not taken by the current crew pairing solution, at a given iteration, even if they are incompatible with those included. For this reason, and to allow the dual variables to take different values, a lifting procedure was proposed which can be used on all types of feasibility cuts. It was shown that when the procedure is applied on a cut corresponding to an extreme ray, the resulting cut also corresponds to an extreme ray. On test instances containing up to 700 daily legs, it was observed that the proposed lifting procedure is useful to strengthen the cuts and to reduce the total computing time.

The paper has also shown that for feasibility cuts with a right-hand-side of value zero, some linear combinations of extreme rays give stronger cuts than extreme rays. This particular type of cuts forbids the use of individually infeasible short connections, and they are generated when some dual variables are only present in the non-negativity constraints. This counter-intuitive result could potentially be applied to situations where some other blocks of dual constraints are independent, or only weakly linked

by the variables. The numerical results have pointed out that the bounding method generating cuts corresponding to extreme rays is preferred only when the individually infeasible short connections are identified in a preprocessing step, *i.e.*, when no zero right-hand-side cuts need to be generated. In a Benders decomposition method, the subproblem formulation does not have to be identical at every iteration. One bounding method could thus be used in the first iterations to generate all the zero right-hand-side cuts, and then, the Benders subproblem could shift to another bounding method.

Acknowledgments

This work was supported by the Québec Government (Fonds pour la Formation de Chercheurs et l'Aide à la Recherche), and the Natural Sciences and Engineering Research Council of Canada. This support is gratefully acknowledged.

Bibliography

- BENDERS, J.F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4, 238–252.
- BERTSIMAS, D. AND TSITSIKLIS, J.N. (1997). *Introduction to Linear Optimization*. Nashua: Athena Scientific.
- COHN, A.M. AND BARNHART, C. (2003). Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research*, 51, 387–396.
- CORDEAU, J.-F., STOJKOVIĆ, G., SOUMIS, F. AND DESROSIERS, J. (2001). Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation Science*, 35, 375–388.

- DESAULNIERS, G., DESROSIERS, J., IOACHIM, I., SOLOMON, M.M., SOUMIS, F. AND VILLENEUVE, D. (1998b). A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. In T.G. Crainic and G. Laporte, eds, *Fleet Management and Logistics*, pages 57–93. Boston: Kluwer.
- HUANG, H.-C., LIANG, Z. AND LI, R. (2003). Solving integrated aircraft routing and crew scheduling problem by Benders decomposition. INFORMS Annual Meeting, Atlanta.
- KLABJAN, D., JOHNSON, E.L. AND NEMHAUSER, G.L. (2002). Airline crew scheduling with time windows and plane count constraints. *Transportation Science*, 36, 337–348.
- MAGNANTI, T.L. AND WONG, R.T. (1981). Accelerating Benders decomposition: Algorithmic enhancement and model selection criteria. *Operations Research*, 29, 464–484.
- MERCIER, A., CORDEAU, J.-F. AND SOUMIS, F. (2005). A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers & Operations Research*, 32, 1451–1476.
- NEMHAUSER, G. L. AND WOLSEY, L. A. (1988). *Integer and Combinatorial Optimization*. New York: Wiley.
- YU, G., ed. (1998). *Operations Research in the Airline Industry*. Boston: Kluwer.

Appendix A

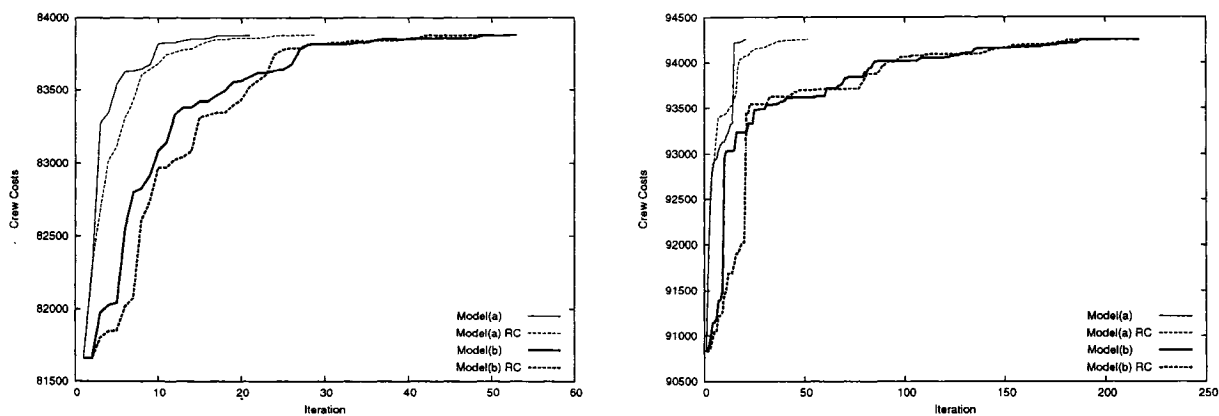


Figure 6.1 Crew costs as a function of the number of iteration - ECP (320D and MD80LA)

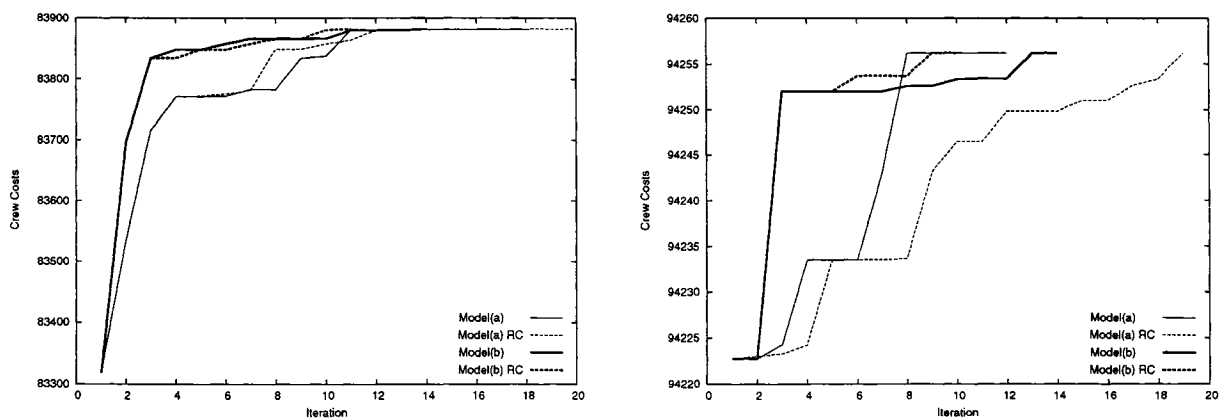


Figure 6.2 Crew costs as a function of the number of iteration - ECP RA (320D and MD80LA)

Appendix B

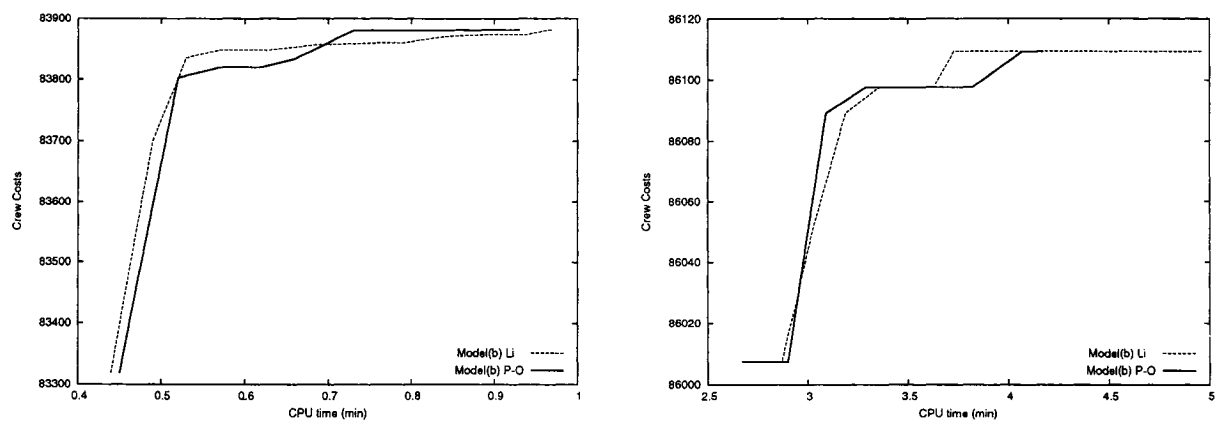


Figure 6.3 Crew costs as a function of CPU time - ECP RA (320D and D9SB)

CHAPITRE 7

DISCUSSION GÉNÉRALE ET CONCLUSION

Comme l'a montré la revue de la littérature, les différents problèmes de la planification tactique en transport aérien ont traditionnellement été résolus de manière séquentielle, et ce, pour en faciliter la résolution. Étant donné le potentiel de gains pour les compagnies aériennes, les travaux récents s'orientent plutôt vers l'intégration, partielle ou complète, des différentes étapes de la planification. La résolution intégrée des problèmes d'itinéraires d'avions et de rotations d'équipages permet en effet de diminuer de façon importante les coûts d'opération des compagnies aériennes. Cette thèse a permis de développer des modèles et des approches de résolution pour traiter des problèmes intégrés d'avions et d'équipages de grande taille. Ce chapitre résume d'abord les principales contributions scientifiques de la thèse. Suit ensuite une discussion des retombées pour l'industrie du transport aérien. Finalement, des suggestions sont données pour des travaux futurs.

7.1 Contributions scientifiques de la thèse

Les contributions de cette thèse se situent à trois niveaux. La première contribution est de présenter deux extensions permettant d'améliorer la qualité des solutions produites par le modèle intégré avions-équipages. Nous avons d'abord présenté une formulation qui, en introduisant un terme de pénalité dans la fonction-objectif pour les connexions dites restreintes, fournit des solutions plus robustes, c'est-à-dire des routes d'avions et des horaires d'équipages moins susceptibles d'être perturbés par des changements imprévus à l'horaire des vols. Nous avons aussi proposé d'améliorer la qualité globale

des solutions en considérant un niveau supplémentaire d'intégration qui permet de prendre certaines décisions par rapport à l'horaire des vols. La formulation compacte qui est proposée pour permettre des modifications à l'horaire des vols produit en effet des solutions à la fois moins chères pour ce qui est des coûts d'équipages et nécessitant un plus petit nombre d'appareils par rapport à un modèle avec un horaire fixe.

L'approche principale proposée dans la littérature pour résoudre le problème intégré d'avions et d'équipages consiste à résoudre, itérativement, un problème de rotations d'équipages modifié, où une coupe de réalisabilité est ajoutée au problème d'équipages jusqu'à ce que l'ensemble des connexions courtes utilisé par les équipages dans la solution courante soit réalisable pour les avions. Les coupes de réalisabilité peuvent être générées par une approche de décomposition de Benders, ou à partir d'une famille prédéfinie de coupes (COHN et BARNHART (2003)). La deuxième contribution de cette thèse a été de montrer théoriquement que les coupes de réalisabilité générées à partir d'une famille prédéfinie sont toutes incluses parmi l'ensemble des coupes de Benders, et peuvent aussi être dominées par ces dernières.

La troisième contribution de cette thèse est l'introduction de raffinements pour améliorer l'efficacité de la méthode de décomposition de Benders pour la résolution du modèle intégré d'avions et d'équipages. Nous avons d'abord proposé d'inverser l'ordre naturel de résolution dans la méthode de décomposition de Benders (l'ordre habituellement choisi par les approches séquentielles), en mettant le problème d'avions dans le sous-problème de Benders. En effet, comme les coûts du problème intégré sont attribuables aux rotations d'équipages, mettre le problème d'avions dans le sous-problème de Benders fait en sorte que celui-ci n'a pas à transférer d'information d'optimalité (de coûts). Par ailleurs, les coupes de Benders peuvent être renforcées en résolvant un problème auxiliaire qui génère des coupes Pareto-optimales. Le problème-maître de Benders étant résolu par génération de colonnes, une représentation explicite du polyèdre n'est pas disponible. Nous avons montré qu'en fixant arbitrairement le

coefficient de certaines variables dans la fonction-objectif du sous-problème auxiliaire dual, il existe malgré tout une façon simple de le résoudre. Lorsque l'approche proposée est comparée, avec les mêmes données, à celle de CORDEAU *et al.* (2001), ces deux raffinements ont permis de réduire les temps de calcul par un facteur de plus de 20, en moyenne, et ce, sans détériorer la qualité des solutions. De plus, cette thèse a montré qu'une formulation plus compacte d'un modèle peut renforcer les coupes de Benders, et que la génération dynamique de contraintes à la fois dans le problème-maître de Benders et dans son sous-problème n'affecte en rien la validité des coupes déjà générées et permet d'accélérer la convergence de la méthode. Il est aussi montré que, dans le cas où l'on veut éviter de devoir identifier des rayons extrêmes, la façon de borner le sous-problème dual de Benders influence grandement l'efficacité de la méthode de résolution de Benders et que certaines combinaisons de rayons extrêmes peuvent même donner des coupes plus fortes que celles données directement par les rayons extrêmes.

En résumé, la décomposition de Benders semble être un bon outil pour résoudre les problèmes intégrés. Cette méthode est flexible et permet de bien résoudre le modèle de base ainsi que certaines extensions de celui-ci. Bien choisir le problème-maître de Benders, utiliser une relaxation de son sous-problème et renforcer les coupes qu'il génère sont quelques exemples qui permettent ensemble de diminuer de façon considérable les temps de calcul nécessaires à la résolution de ces problèmes intégrés. Lorsqu'elle est comparée à notre implantation de la méthode proposée par COHN et BARNHART (2003), l'approche proposée dans cette thèse a trouvé des solutions qui sont à la fois plus robustes et moins coûteuses, et ce, en réduisant les temps de calcul.

7.2 Retombées pour l'industrie du transport aérien

Cette thèse a montré qu'il est possible de résoudre des problèmes tactiques intégrés comprenant jusqu'à 500 vols quotidiens lorsque des modifications à l'horaire sont permises, et jusqu'à 700 vols lorsque l'horaire est fixe. Il est important de noter que les résultats présentés dans cette thèse ont été obtenus à l'aide d'ordinateurs dont la puissance est moins élevée que les ordinateurs les plus puissants disponibles à ce jour. Ceci porte à croire que l'approche proposée pourrait résoudre, avec des temps de calcul raisonnables, le problème intégré pour les plus grosses flottes d'avions des transporteurs aériens actuels (environ 1000 vols par jour pour un même type d'avion), ou du moins le pourrait prochainement.

De nos jours, les transporteurs aériens résolvent les problèmes de la planification tactique de la façon suivante. L'affectation des types d'appareils, sur une semaine, est premièrement fait en construisant en même temps des itinéraires d'avions, mais sans inclure les contraintes d'entretien (étape A). Des itinéraires d'avions respectant les contraintes d'entretien sont ensuite créés (étape B), pour finalement constituer les rotations d'équipages (étape C) qui seront combinées pour obtenir les horaires mensuels des équipages (étape D).

À court ou à moyen terme, le problème intégré d'itinéraires d'avions et de rotations d'équipages quotidien qui est résolu dans cette thèse pourrait s'insérer dans le processus de planification à l'étape B, de manière à l'enrichir. Les grandes compagnies aériennes résolvent d'ailleurs généralement les problèmes d'itinéraires d'avions et de rotations d'équipages sur une journée, avant de traiter les exceptions. Un meilleur ensemble de connexions courtes serait ainsi choisi, et celui-ci pourrait ensuite être utilisé directement pour fabriquer, à l'étape C, des rotations d'équipages de meilleure qualité. Ce dernier problème doit être résolu, car le problème intégré utilise une

structure de coûts simplifié. De plus en plus, les transporteurs aériens choisissent de modifier graduellement l'horaire des vols entre deux saisons. Ceci donne davantage d'importance à l'étape B qui, contrairement à l'étape A, doit être résolue pour chacune des modifications à l'horaire.

Contrairement aux vols offerts par les grandes compagnies aériennes, ceux qui sont offerts par les transporteurs de petite ou moyenne taille sont plutôt irréguliers, et ces derniers doivent, par conséquent, résoudre les problèmes d'itinéraires d'avions et de rotations d'équipages sur une semaine. L'approche proposée dans cette thèse pourrait néanmoins s'insérer dans le processus de planification. Cette fois-ci, le problème intégré quotidien serait résolu avant l'étape B, pour choisir un bon ensemble de connexions courtes. L'étape B serait ensuite résolue pour fabriquer, sur une semaine, des itinéraires d'avions qui utilisent, autant que possible, l'ensemble de connexions courtes choisi par le modèle intégré. Ceci pourrait être fait en pénalisant les autres connexions courtes. Le coût des rotations d'équipages construites à l'étape C en serait, par le fait même, diminué.

7.3 Travaux futurs

Avec les gains de puissance des ordinateurs qui sont à prévoir dans les prochaines années, la popularité des méthodes intégrées sera en croissance. L'intégration de l'affectation des types d'appareils (l'étape A) au problème de la création des itinéraires d'avions et des rotations d'équipages (les étapes B et C) permettra à coup sûr des diminutions impressionnantes de coûts pour les transporteurs aériens.

La difficulté principale de ce dernier problème provient du très grand nombre de contraintes liant les trois étapes. Ces contraintes sont nécessaires afin de s'assurer que l'équipage affecté à un vol possède des qualifications compatibles avec le type

d'appareil affecté à ce même vol. Ce problème intégré possède une structure appropriée pour une décomposition de Benders. Le problème-maître de Benders pourrait traiter le problème des avions en entier, c'est-à-dire le choix des types d'appareils et la création des itinéraires d'avions et le sous-problème pourrait déterminer les rotations d'équipages. Lorsque seraient fixés les types d'appareils pour chacun des segments de vol par le problème-maître, le problème se séparerait en un sous-problème d'équipage par type d'appareil, ou par famille d'appareils. Les différents sous-problèmes pourraient donc être résolus en parallèle, ce qui serait impossible sans décomposition.

Le développement de nouvelles méthodes de séparation et d'évaluation progressive pourrait aussi accélérer la méthode de résolution, et ce, peu importe le degré d'intégration des modèles. Il existe un grand nombre de possibilités pour le choix des contraintes à ajouter à chaque noeud de branchement. Il est important de bien les choisir : on peut évidemment fixer des chemins dans le réseau, mais on pourrait aussi privilégier certains types d'avions plus coûteux au début, forcer certaines connexions à des aéroports secondaires peu fréquentés ou utiliser des variables d'horaire, s'il y a lieu. La fonction-objectif du problème intégré qui inclut aussi l'affectation des types d'appareils comporte à la fois des variables d'équipages et des variables d'avions (la création des itinéraires d'avions n'était quant à elle qu'un problème de réalisabilité). L'écart d'intégrité pourrait être plus grand pour ce problème, et le choix des contraintes à ajouter à chaque noeud de branchement devient donc primordial.

Bibliographie

- ABARA, J. (1989). Applying integer linear programming to the fleet assignment problem. *Interfaces*, 19, 211–232.
- AHUJA, R.K., LIU, J., ORLIN, J.B., GOODSTEIN, J. ET MUKHERJEE, A. (2004). A neighborhood search algorithm for the combined through and fleet assignment model with time windows. *Networks*, 44, 160–171.
- AHUJA, R.K. ET ORLIN, J.B. (2002). Very large-scale neighborhood search for airline fleet scheduling. *SIAM News*, 35, 1–4.
- ANBIL, R., TANGA, R. ET JOHNSON, E.L. (1992). A global approach to crew-pairing optimization. *IBM Systems Journal*, 31, 71–78.
- ATAC. *Transport Canada Assumptions Report 2005-2019*. (2005). Air Transport Association of Canada, <http://www.atac.ca>.
- BALL, M., BODIN, L. ET DIAL, R. (1983). A matching based heuristic for scheduling mass transit crews and vehicles. *Transportation Science*, 17, 4–31.
- BARNHART, C., BOLAND, N.L., CLARKE, L.W., JOHNSON, E.L., NEMHAUSER, G.L. ET SHENOI, R.G. (1998a). Flight string models for aircraft fleet and routing. *Transportation Science*, 32, 208–220.
- BARNHART, C., COHN, A.M., JOHNSON, E.L., KLABJAN, D., NEMHAUSER, G.L. ET VANCE, P.H. (2003). Airline crew scheduling. In R.W. Hall, rédacteur, *Handbook of Transportation Science (2e édition)*, pages 517–560. Boston : Kluwer.
- BARNHART, C., HATAY, L. ET JOHNSON, E.L. (1995). Deadhead selection for the long-haul crew pairing problem. *Operations Research*, 43, 491–499.

- BARNHART, C., JOHNSON, E.L., NEMHAUSER, G.L., SAVELSBERGH, M.W.P. ET VANCE, P.H. (1998b). Branch-and-price : Column generation for solving huge integer programs. *Operations Research*, 46, 316–329.
- BARNHART, C., LU, F. ET SHENOI, R. (1998c). Integrated airline schedule planning. In *Operations Research in the Airline Industry*, pages 384–403. Boston : Kluwer.
- BARNHART, C. ET SHENOI, R.G. (1998). An approximate model and solution approach for the long-haul crew pairing problem. *Transportation Science*, 32, 221–231.
- BELOBABA, P.P. ET FARKAS, A. (1999). Yield management impacts on airline spill estimation. *Transportation Science*, 33, 217–232.
- BENDERS, J.F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4, 238–252.
- BERTSIMAS, D. ET TSITSIKLIS, J.N. (1997). *Introduction to Linear Optimization*. Nashua : Athena Scientific.
- CHEBALOV, S. ET KLABJAN, D. (2006). Robust airline crew scheduling : Move-up crews. *Transportation Science*. À paraître.
- CHU, H.D., GELMAN, E. ET JOHNSON, E.L. (1997). Solving large scale crew scheduling problems. *European Journal of Operational Research*, 97, 260–268.
- CLARKE, L.W., HANE, C.A., JOHNSON, E.L. ET NEMHAUSER, G.L. (1996). Maintenance and crew considerations in fleet assignment. *Transportation Science*, 30, 249–260.
- CLARKE, L.W., JOHNSON, E.L., NEMHAUSER, G.L. ET ZHU, Z. (1997). The aircraft rotation problem. *Annals of Operations Research*, 69, 33–46.

- COHN, A.M. ET BARNHART, C. (2003). Improving crew scheduling by incorporating key maintenance routing decisions. *Operations Research*, 51, 387–396.
- CORDEAU, J.-F., STOJKOVIĆ, G., SOUMIS, F. ET DESROSIERS, J. (2001). Benders decomposition for simultaneous aircraft routing and crew scheduling. *Transportation Science*, 35, 375–388.
- DASKIN, M.S. ET PANAYOTOPOULOS, N.D. (1989). A Lagrangian relaxation approach to assigning aircraft to routes in hub and spoke network. *Transportation Science*, 23, 91–99.
- DESAULNIERS, G., DESROSIERS, J., DUMAS, Y., MARC, S., RIOUX, B., SOLOMON, M.M. ET SOUMIS, F. (1997a). Crew pairing at Air France. *European Journal of Operational Research*, 97, 245–259.
- DESAULNIERS, G., DESROSIERS, J., DUMAS, Y., SOLOMON, M.M. ET SOUMIS, F. (1997b). Daily aircraft routing and scheduling. *Management Science*, 43, 841–855.
- DESAULNIERS, G., DESROSIERS, J., GAMACHE, M. ET SOUMIS, F. (1998a). Crew scheduling in air transportation. In T. G. Crainic et G. Laporte, rédacteurs, *Fleet Management and Logistics*, pages 169–185. Boston : Kluwer.
- DESAULNIERS, G., DESROSIERS, J., IOACHIM, I., SOLOMON, M.M., SOUMIS, F. ET VILLENEUVE, D. (1998b). A unified framework for deterministic time constrained vehicle routing and crew scheduling problems. In T.G. Crainic et G. Laporte, rédacteurs, *Fleet Management and Logistics*, pages 57–93. Boston : Kluwer.
- EHRGOTT, M. ET RYAN, D.M. (2002). Constructing robust crew schedules with bicriteria optimization. *Journal of Multi-Criteria Decision Analysis*, 11, 139–150.
- FEO, T.A. ET BARD, J.F. (1989). Flight scheduling and maintenance base planning. *Management Science*, 35, 1415–1432.

- FRELING, R. (1997). *Models and Techniques for Integrating Vehicle and Crew Scheduling*. Thèse de doctorat, Tinbergen Institute, Atlanta.
- FRELING, R., BOENDER, G. ET PAIXAO, A. (1995). An integrated approach to vehicle and crew scheduling. Rapport technique 9503/A, Erasmus University, Rotterdam.
- GOPALAN, R. ET TALLURI, K.T. (1998). The aircraft maintenance routing problem. *Operations Research*, 46, 260–271.
- GRAVES, G.W., MCBRIDE, R.D., GERSHKOFF, I., ANDERSON, D. ET MAHIDHARA, D. (1993). Flight crew scheduling. *Management Science*, 39, 736–745.
- GU, Z., JOHNSON, E.L., NEMHAUSER, G.L. ET WANG, Y. (1994). Some properties of the fleet assignment problem. *Operations Research Letters*, 15, 59–71.
- HAASE, K., DESAULNIERS, G. ET DESROSIERS, J. (2001). Simultaneous vehicle and crew scheduling in urban mass transit systems. *Transportation Science*, 35, 286–303.
- HAASE, K. ET FRIBERG, C. (1999). An exact branch and cut algorithm for the vehicle and crew scheduling problem. In *Computer-Aided Transit Scheduling*, pages 63–80. N.H.M. Wilson, Berlin : Springer.
- HANE, C.A., BARNHART, C., JOHNSON, E.L., MARSTEN, R.E., NEMHAUSER, G.L. ET SIGISMONDI, G. (1995). The fleet assignment problem : Solving a large-scale integer program. *Mathematical Programming*, 70, 211–232.
- HOFFMAN, K.L. ET PADBERG, M. (1993). Solving airline crew scheduling problems by branch-and-cut. *Management Science*, 39, 657–682.
- HOUSOS, E. ET ELMROTH, T. (1997). Automatic optimization of subproblems in scheduling airline crews. *Interfaces*, 27, 68–77.

- HUANG, H.-C., LIANG, Z. ET LI, R. (2003). Solving integrated aircraft routing and crew scheduling problem by Benders decomposition. INFORMS Annual Meeting, Atlanta.
- HUISMAN, D., FRELING, R. ET WAGELMANS, A.P.M. (2005). Multiple-depot integrated vehicle and crew scheduling. *Transportation Science*, 39, 491–502.
- IOACHIM, I., DESROSIERS, J., SOUMIS, F. ET BÉLANGER, N. (1999). Fleet assignment and routing with schedule synchronization constraints. *European Journal of Operational Research*, 119, 75–90.
- JARRAH, A.I. ET STREHLER, J.C. (2000). An optimisation model for assigning through flights. *IIE Transactions*, 32, 237–244.
- KESKINOCAK, P. ET TAYUR, S. (1998). Scheduling of time-shared jet aircraft. *Transportation Science*, 32, 277–294.
- KLABJAN, D. (2005). Large-scale models in the airline industry. In M.M. Solomon G. Desaulniers, J. Desrosiers, rédacteurs, *Column Generation*, pages 163–195. New York : Springer.
- KLABJAN, D., JOHNSON, E.L. ET NEMHAUSER, G.L. (2002). Airline crew scheduling with time windows and plane count constraints. *Transportation Science*, 36, 337–348.
- LAVOIE, S., MINOUX, M. ET ODIER, E. (1988). A new approach for crew pairing problems by column generation with an application to air transportation. *European Journal of Operational Research*, 35, 45–58.
- LEVIN, A. (1971). Scheduling and fleet routing models for transportation systems. *Transportation Science*, 5, 232–255.

- MAGNANTI, T.L. ET WONG, R.T. (1981). Accelerating Benders decomposition : Algorithmic enhancement and model selection criteria. *Operations Research*, 29, 464–484.
- MERCIER, A., CORDEAU, J.-F. ET SOUMIS, F. (2005). A computational study of Benders decomposition for the integrated aircraft routing and crew scheduling problem. *Computers & Operations Research*, 32, 1451–1476.
- NEMHAUSER, G. L. ET WOLSEY, L. A. (1988). *Integer and Combinatorial Optimization*. New York : Wiley.
- PATRIKALAKIS, I. ET XEROCOSTAS, D. (1992). A new decomposition scheme of the urban public transport scheduling problem. In M. Desrochers et J.M. Rousseau, rédacteurs, *Computer-Aided Transit Scheduling*, pages 407–425. Berlin : Springer-Verlag.
- REXING, B., BARNHART, C. ET KNIKER, T. (2000). Airline fleet assignment with time windows. *Transportation Science*, 34, 1–20.
- RICHARDSON, R. (1976). An optimization approach to routing aircraft. *Transportation Science*, 10, 52–71.
- ROSENBERG, J.M., JOHNSON, E.L. ET NEMHAUSER, G.L. (2004). A robust fleet assignment model with hub isolation and short cycles. *Transportation Science*, 38, 357–368.
- RUSHMEIER, R.A. ET KONTOGIORGIS, S.A. (1997). Advances in the optimization of airline fleet assignment. *Transportation Science*, 31, 159–169.
- SUBRAMANIAN, R., JR., R.P. SCHEFF, QUILLINAN, J.D., WIPER, D.S. ET MARSTEN, R.E. (1994). Coldstart : Fleet assignment at Delta air lines. *Interfaces*, 24, 104–120.

- TALLURI, K.T. (1996). Swapping applications in a daily airline fleet assignment. *Transportation Science*, 30, 237–248.
- TALLURI, K.T. (1998). The four-day aircraft maintenance routing problem. *Transportation Science*, 32, 43–53.
- TOSINI, E. ET VERCELLIS, C. (1992). An interactive system for extra-urban vehicle and crew scheduling problems. In J.R. Daduna et A. Wren, rédacteurs, *Computer-Aided Transit Scheduling*, pages 41–53. Berlin : Springer-Verlag.
- VANCE, P.H., BARNHART, C., JOHNSON, E.L. ET NEMHAUSER, G.L. (1997). Airline crew scheduling : A new formulation and decomposition algorithm. *Operations Research*, 45, 188–200.
- YU, G., rédacteur. (1998). *Operations Research in the Airline Industry*. Boston : Kluwer.