



HAL
open science

Modèles de dimensionnement et de planification dans un centre d'appels

Rabie Nait-Abdallah

► **To cite this version:**

Rabie Nait-Abdallah. Modèles de dimensionnement et de planification dans un centre d'appels. Sciences de l'ingénieur [physics]. Ecole Centrale Paris, 2008. Français. NNT : . tel-00275832

HAL Id: tel-00275832

<https://theses.hal.science/tel-00275832>

Submitted on 25 Apr 2008

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

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



**Thèse de doctorat de l'Ecole Centrale des Arts et Manufactures
Spécialité : Génie Industriel**

Présentée par
M. Rabie NAIT-ABDALLAH

Le 18 janvier 2008
Pour obtenir le grade de Docteur de l'Ecole Centrale Paris

Sujet de thèse :

**MODÈLES DE DIMENSIONNEMENT ET DE
PLANIFICATION DANS UN CENTRE D'APPELS**

Jury

Président	E. JACQUET-LAGRÈZE, EURODÉCISION
Rapporteurs	M. GOURGAND, Professeur - ISIMA G. KOOLE, Professeur - VU University Amsterdam
Directeurs de thèse	F. CHAUVET, Gaz de France Y. DALLERY, Professeur – Ecole Centrale de Paris

Laboratoire Génie Industriel
Ecole Centrale Paris
Grande Voie des Vignes
92925 Châtenay-Malabry Cedex

N° 2008-03

REMERCIEMENTS

Je tiens à remercier Fabrice Chauvet qui a encadré mes recherches et qui m'a fait confiance il y a bien longtemps en me recrutant au sein de son équipe d'optimisation. Ce travail n'aurait pas abouti sans son suivi et son implication. Une grande partie des idées mises en œuvre dans cette thèse ont été le fruit de ses réflexions.

Ma gratitude va également à Yves Dallery qui a dirigé mes travaux. Sa hauteur de vue et ses intuitions toujours pertinentes ont beaucoup enrichi cette thèse.

Je remercie Michel Gourgand et Ger Koole pour l'intérêt qu'ils ont porté à cette thèse en acceptant d'en être les rapporteurs. Ma reconnaissance va également à Eric Jacquet-Lagrèze qui me fait l'honneur d'être le président de mon jury.

Mes remerciements vont aussi à Ali Cheaitou qui m'a apporté son aide pour les recherches bibliographiques.

Merci enfin à tous ceux, famille et amis, qui m'ont encouragé et soutenu. Je ne pourrais tous les citer sous peine d'en oublier.

TABLE DES MATIÈRES

Introduction.....	4
Chapitre I	5
Introduction et Problématique La gestion des ressources humaines, un problème central dans un centre d'appels	5
I.1 Introduction à l'univers des centres d'appels	6
I.1.1 Quelques notions de base concernant les centres d'appels.....	6
I.1.1.1 Définition d'un centre d'appels.....	6
I.1.1.2 Fonctionnement d'un centre d'appels	7
I.1.2 Les horizons de décision dans un centre d'appels	8
I.1.2.1 Décisions stratégiques : Investissements et choix structurants	8
I.1.2.2 Décisions tactiques : gestion des ressources humaines	9
I.1.2.3 Décisions opérationnelles et temps réel : Pilotage temps réel de l'activité.....	10
I.1.3 Focus sur l'horizon tactique : la chaîne de décision en gestion des ressources humaines	12
I.2 Enjeux du <i>shift-scheduling</i> , un problème central dans l'optimisation des centres d'appels	15
I.2.1 Enjeu industriel : un problème au cœur de l'adéquation entre charge et capacité	15
I.2.1.1 Une charge d'appels à forte variabilité	15
I.2.1.2 Une obligation de réponse en temps réel	16
I.2.2 Enjeu scientifique : un modèle d'optimisation de plusieurs problèmes de gestion des ressources humaines	18
I.3 Problématiques : les axes de contribution de la thèse	19
I.3.1 L'incomplétude de l'information, un principe au fondement de l'organisation des décisions ..	19
I.3.2 Proposer un paradigme et un modèle unificateur du <i>shift-scheduling</i>	20
I.3.3 Prendre en compte la qualité de service de manière explicite	20
Chapitre II.....	23
Organisation des décisions de gestion des ressources humaines.....	23
II.1 Méthodologie d'organisation des décisions de gestion des ressources humaines dans les centres d'appels 24	
II.1.1 Cas déterministe avec une décision centralisée	25
II.1.2 Le principe de décomposition - impact de l'incertitude	26
II.1.2.1 Décomposition dimensionnement-planification	28
II.1.2.2 Décomposition de la planification journalière et horaire	29
II.1.3 Le principe de décomposition – impact du partage de la décision	30
II.1.4 Limites de la décomposition des décisions	32
II.1.5 Introduction des décisions d'ajustement de capacité	34
II.1.6 Synthèse.....	36
II.2 Etat de l'art des problèmes de gestion des ressources humaines dans les services	39
Chapitre III	43
Dimensionnement sous contrainte de couverture	43
III.1 Position du problème.....	44
III.2 Etat de l'art des modélisations du problème de construction de vacations (<i>shift scheduling problem</i>)	47
III.2.1 Définition du problème de construction de vacations.....	47
III.2.1.1 Quelques définitions :	47
III.2.2 Les premiers modèles	48
III.2.2.1 Programme linéaire en nombres entiers (Dantzig 1954).....	48

III.2.2.2	Problème de flot (Segal 1974).....	49
III.2.3	Les modèles implicites.....	50
III.2.3.1	Le concept de modélisation implicite	50
III.2.3.2	Début, fin et durée de vacations implicites (Moondra 1976)	50
III.2.3.3	Une seule pause implicite (Bechtold et Jacobs 1990)	51
III.2.3.4	Modèle doublement implicite (Thompson 1995).....	52
III.2.3.5	Plusieurs pauses implicites (Aykin 1996)	53
III.2.4	Synthèse des modèles	53
III.3	Modélisation via le paradigme de chaîne d'activité	54
III.3.1	Le paradigme de chaîne d'activité	54
III.3.2	Modélisation mathématique.....	57
III.3.2.1	Définitions et notations	57
III.3.3	Exemple d'application et limites du paradigme.....	59
III.3.3.1	Illustration du paradigme	59
III.3.3.2	Limites du paradigme de chaîne d'activités	60
III.3.4	Les variables de décision	61
III.3.5	Les contraintes	62
III.3.5.1	Contraintes de temps de travail	62
III.3.5.2	Contraintes de couverture	65
III.3.6	Fonction objectif - Minimisation du coût	65
III.3.7	Récapitulatif du modèle	66
III.4	Analyse des propriétés du modèle proposé	68
III.4.1	Equivalence entre solution implicite et solution individuelle	68
III.4.2	Analyse de la difficulté du problème.....	71
III.4.2.1	La totale unimodularité de la matrice des contraintes de temps de travail.....	72
III.4.2.2	Introduction des contraintes de couverture	77
III.4.2.3	Cas particulier polynomial	78
III.5	Généralisation du modèle issu du paradigme de chaîne d'activité à la multi-compétence.....	79
III.5.1	Plusieurs flux d'appels (synchrone).....	79
III.5.2	Plusieurs flux stockable (asynchrone) et plusieurs flux d'appels	80
III.5.2.1	Un seul flux stockable.....	81
III.5.2.2	Plusieurs flux stockable	82
III.5.3	Le cas du <i>Call blending</i>	83
III.5.4	Equivalence implicite-explicite du modèle généralisé.....	85
III.6	Positionnement du modèle par rapport à la littérature	86
III.7	Résolution du problème de dimensionnement - Qualité de la relaxation linéaire	87
III.7.1	Faisabilité de la solution arrondie.....	87
III.7.2	Qualité de la solution arrondie.....	88
III.8	Conclusion.....	92

Chapitre IV.....95

Planification maximisant la qualité de service	95	
IV.1	Position du problème.....	96
IV.2	Prise en compte de la qualité de service.....	98
IV.2.1	Qualité de service à la journée plutôt que par période	98
IV.2.2	Impact de la non-linéarité de la qualité de service sur le shift scheduling	98
IV.2.3	Etat de l'art	99
IV.2.3.1	Couplage par la fonction objectif : qualité de service globale à la journée.....	99
IV.2.3.2	Couplage par influence mutuelle entre périodes	100
IV.3	Modélisation.....	102
IV.3.1	Hypothèses.....	102
IV.3.2	Contraintes	102
IV.3.3	Fonction objectif – optimisation de la qualité de service.....	103
IV.3.3.1	Deux approches de modélisation de la qualité de service	104
IV.3.4	Synthèse du modèle	109
IV.3.4.1	Contraintes de temps de travail	109
IV.3.4.2	Contraintes de limitation des effectifs.....	109
IV.3.4.3	Fonction objectif	109
IV.4	Comparaison entre la vision déterministe et la vision stochastique	111
IV.5	Conclusion.....	115

Conclusions et perspectives..... 117
Bibliographie 121

INTRODUCTION

Les centres d'appels sont aujourd'hui incontournables dans les activités de la vie quotidienne. Ils sont le moyen de fournir toutes sortes de services : centrales de réservation, vente par correspondance, assistance technique ou renseignements, etc. Toutes ces activités sont assurées à distance par des agents ou conseillers de clientèle. Aussi, la problématique de gestion des ressources humaines est centrale car d'une part, elle conditionne la qualité du service rendu et d'autre part, elle représente l'élément de coût le plus significatif.

Cette thèse aborde cette problématique de gestion des ressources humaines dans un centre d'appels. Elle est construite autour de quatre chapitres :

Le premier chapitre présente la problématique et les axes de recherche que nous avons identifiés pour cette thèse.

Dans le second chapitre, nous analysons les fondements et mécanismes qui conditionnent l'organisation des décisions dans un centre d'appels.

Dans le troisième chapitre, nous abordons le problème du dimensionnement d'un centre d'appels. Nous développons également notre concept de paradigme de chaîne d'activités.

Dans le quatrième chapitre, nous nous intéressons au problème de planification et nous présentons nos travaux sur la modélisation de la qualité de service.

Chapitre I

INTRODUCTION ET PROBLÉMATIQUE LA GESTION DES RESSOURCES HUMAINES, UN PROBLÈME CENTRAL DANS UN CENTRE D'APPELS

L'objectif de ce chapitre est d'une part de présenter les centres d'appels et le contexte les entourant et d'autre part de positionner les contributions de la thèse qui structureront la suite du document.

Nous commençons par présenter les enjeux économiques et industriels des centres d'appels et nous donnons un rapide aperçu des décisions qu'implique ce type d'activité

Par la suite, nous nous focalisons sur le problème qui est au cœur de cette thèse : le problème de construction de vacations ou *shift-scheduling*. Il présente un enjeu central à la fois d'un point de vue industriel et scientifique.

Nous concluons enfin par une description de la problématique et des thèmes qui seront développés dans la suite de la thèse.

I.1 Introduction à l'univers des centres d'appels

I.1.1 Quelques notions de base concernant les centres d'appels

I.1.1.1 Définition d'un centre d'appels

Un centre d'appels a pour finalité la gestion de la relation que les entreprises souhaitent entretenir avec leurs clients. Il est l'objet de plusieurs enjeux d'importance pour l'entreprise : un enjeu d'image lié à la qualité de service perçue par le client, un enjeu financier, du fait de la masse salariale que représentent les conseillers et un enjeu social relatif aux conditions de travail.

Du fait de leur rôle central dans la constitution de la relation client, les centres d'appels se retrouvent dans pratiquement tous les secteurs d'activités : banque, assurance, téléphonie, vente à distance, services administratifs, etc. La Figure I-1 donne un aperçu de l'environnement de travail dans un centre d'appels (source : centre d'appels de la banque *First Direct* (Gans, Koole et al. 2003))



Figure I-1 : Environnement de travail du centre d'appels de la banque *First Direct* (Gans, Koole et al. 2003)

Une définition des centres d'appels plus détaillée est donnée dans (Jacquinet 2007) :

Les centres d'appels dont la vocation est de gérer à distance la relation que les entreprises souhaitent entretenir avec leurs clients et prospects : c'est un ensemble de moyens humains, organisationnels et techniques mis en place afin d'apporter à la demande de chaque client une réponse adaptée. A ce titre, les centres d'appels se définissent comme des entités composées d'opérateurs organisés par type de compétences et regroupés par équipe sur des plate-formes destinées à gérer, exclusivement par téléphone, des clients et/ou des prospects en s'appuyant sur des systèmes de couplage téléphonique et informatique, que ce soit en émission ou en réception d'appels. Entités de relation à distance, les centres d'appels optimisent l'outil téléphonique et ses connexions avec l'informatique et d'autres médias (courrier, fax, minitel, internet, extranet, sms, wap, etc.). Ils mettent en jeu quatre composantes majeures :

- *Les ressources humaines (téléconseillers, superviseurs, manager, formateurs ...)*
- *La technologie (téléphonie, informatique, internet, logiciels, progiciels, serveurs multimédia, bases de données, cartes de communication, câblage ...)*
- *La logistique (immobilier, mobilier, ergonomie de l'environnement matériel et de l'environnement 'écran')*
- *Une culture et des méthodes marketing (stratégie de l'entreprise, relation client, fulfillment, profitabilité...)*

Source : convention collective prestataires de services dans le domaine du secteur tertiaire supplément n°10 Brochure 3310 – 2^e édition – décembre 2000 – avenant du 18 septembre 2001 relatif à la modification du champ d'application

I.1.1.2 Fonctionnement d'un centre d'appels

Le fonctionnement d'un centre d'appels est illustré dans la Figure I-2. Les appels arrivent dans une file et sont mis en attente jusqu'à ce qu'un conseiller de clientèle soit disponible pour leur répondre. L'appel est alors *répondu* ou *servi*. La taille de cette file d'attente peut être limitée. Lorsque cette limite est atteinte, les appels qui arrivent ne peuvent plus être pris en charge. Généralement, un message les invite à rappeler ultérieurement. On parle alors d'appels *déconnectés* ou *dissuadés*. Par ailleurs, tous les appels mis en file d'attente ne seront pas nécessairement servis. En effet, certains clients peuvent s'impatisser et décider de quitter la file d'attente. On parle alors d'appels *abandonnés*. Parmi les appels déconnectés et abandonnés, une partie renouvelle son appel. Ceci constitue un flux de *rappels* qui s'additionne au flux d'appels originel pour constituer la totalité des appels *reçus*.

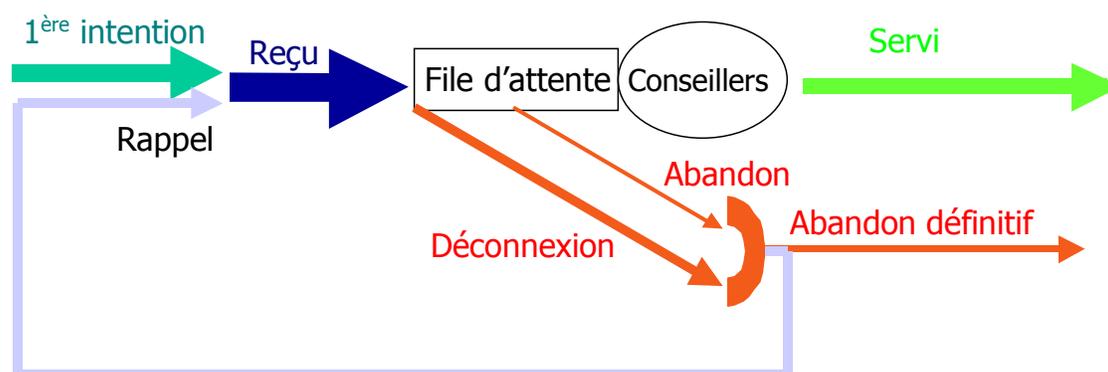


Figure I-2 : Schéma représentatif du fonctionnement d'un centre d'appels

I.1.2 Les horizons de décision dans un centre d'appels

L'objet de cette partie est de décrire les différents horizons de décision liés au management des centres d'appels. Nous distinguons trois horizons de décision : le niveau stratégique correspondant aux décisions les plus structurantes, le niveau tactique lié aux problématiques de gestion des ressources humaines et le niveau opérationnel relatif au pilotage de l'activité au cours de la journée et en temps réel. Les principales décisions de ces horizons de décision sont synthétisées dans le Tableau I-1.

I.1.2.1 Décisions stratégiques : Investissements et choix structurants

Les décisions à ce niveau ont pour finalité la mise en place d'un centre d'appels qui corresponde aux objectifs de l'entreprise. Ces décisions relèvent principalement de deux aspects :

a) Infrastructures et technologies

Il s'agit de déterminer l'emplacement géographique du site (ou des sites dans le cas d'un centre d'appels multi-site), la nature des technologies informatiques et télécoms à mettre en place ainsi que les fournisseurs de solutions (Postes informatiques et téléphoniques, Progiciels de gestion de la relation client, Outils de pilotage et de supervision, etc.).

b) Organisation

L'un des aspects les plus fondamentaux en termes d'organisation est le positionnement du centre d'appels, et plus généralement de la relation client, au sein de l'entreprise.

En effet, le centre d'appels peut être un canal de contact avec le client parmi d'autres sans être le cœur de la relation avec les clients. Dans ce cas, il est généralement considéré comme centre de coût nécessaire qui doit être minimisé. L'entreprise sera alors d'autant plus disposée à l'externaliser s'il correspond à une fonction qui n'est pas centrale dans son activité. Dans d'autres cas, le centre d'appels est au cœur de la politique d'acquisition et de fidélisation des clients des entreprises. Ce n'est plus un centre de coût mais un centre de profit qui crée de la valeur pour l'entreprise.

D'autres décisions majeures doivent également être prises. Notamment :

- Les indicateurs de performance permettant de piloter l'activité : temps moyen d'attente, indicateurs qualitatifs de satisfaction client, volumes de ventes réalisés via le centre d'appels, etc.
- Choix d'externalisation ou d'internalisation du centre d'appels.
- Nature des flux traités et canaux d'origine : appels entrants, appels sortants, emails, courriers, fax, etc.
- Profil de compétence des conseillers : polyvalents ou spécialisés.
- Conditions de travail : nature des contrats de temps de travail (CDI, CDD, temps plein ou temps partiel), formations, plans de carrières au sein de l'entreprise, etc.

Ces choix organisationnels sont très structurants dans la mesure où ils constituent un socle de contraintes. Ils déterminent les marges de manœuvre de toutes les décisions tactiques et opérationnelles qui seront détaillées par la suite. De ce fait, un des enjeux majeurs des décisions à long terme est de « préparer le terrain » pour les décisions à moyen et court terme. Il s'agira par exemple, de choisir judicieusement l'organisation à mettre en place afin de garantir à la fois qualité de service et réactivité (une illustration de l'impact des choix organisationnels sur la performance d'un centre d'appels pourra être trouvée dans (Jouini, Dallery et al. 2004; Jouini, Dallery et al. 2007). On peut citer également l'importance de la politique RH (équilibre CDD/CDI, contrats de travail, plan d'évolution de carrière) qui doit à la fois permettre suffisamment de souplesse pour une planification efficace et satisfaire les conseillers de clientèle afin de les fidéliser et réduire le turnover.

I.1.2.2 Décisions tactiques : gestion des ressources humaines

L'essentiel des décisions à un horizon tactique concerne la gestion des ressources humaines. Nous pouvons les classer dans trois catégories :

a) Le dimensionnement des ressources humaines

Le dimensionnement des ressources humaines consiste à déterminer l'effectif à recruter. La nature des recrutements (CDI, CDD, temps plein ou temps partiel) qui en découlent dépendra de la politique RH de l'entreprise et des contraintes opérationnelles.

Notons que même si les contrats flexibles (CDD, Intérimaires, etc.) permettent une plus grande réactivité, de plus en plus d'entreprises favorisent le recours à des contrats à long terme. C'est un élément clé de fidélisation de leurs conseillers. En effet, compte tenu des coûts de recrutement et de formation, il est dans l'intérêt de l'entreprise de garder les mêmes effectifs aussi longtemps que possible. Ceci est d'autant plus vrai que dans le domaine des centres d'appels les conseillers expérimentés jouent un rôle important dans la formation des nouveaux arrivants (voir à ce sujet (Gaimon et Thompson 1984)).

b) La planification

Le dimensionnement a pour finalité de déterminer le nombre de conseillers à recruter. L'objectif de la planification est d'organiser leur temps de travail afin de respecter au mieux les contraintes de service et leurs contraintes personnelles. Concrètement, la planification en entreprise se décline en un ensemble de processus et d'outils visant à déterminer les jours de congé, les jours de repos ou les emplois du temps par exemple.

c) Ajustement de la capacité de service

Les phases de dimensionnement et de planification peuvent ne pas être suffisantes pour garantir l'adéquation entre charge et capacité. En effet, des situations imprévues de surcharge peuvent conduire l'entreprise à réajuster de manière réactive la capacité de service. Ces besoins de main d'œuvres supplémentaire peuvent être satisfaites par exemple par des recrutements ponctuels ou des prestataires externes.

I.1.2.3 Décisions opérationnelles et temps réel : Pilotage temps réel de l'activité

a) Gestion opérationnelle des imprévus

A l'horizon opérationnel, l'emploi de temps des conseillers est totalement défini. Toutefois, on se réserve généralement une marge de manœuvre afin de gérer les imprévus (surcharge d'appels par rapport à la prévision, incident technique, etc.). Ainsi, en cas de surcharge d'appels par exemple, le management peut décider de manière réactive d'affecter des conseillers à d'autres activités ou de solliciter du personnel en renfort d'autres services.

b) Distribution temps réel des flux aux conseillers

La distribution des flux aux conseillers se fait par les équipements informatiques et télécoms du centre d'appels. Des algorithmes spécifiques permettent d'acheminer les appels vers le conseiller le plus adéquats tout en minimisant le temps d'attente. Des distinctions en termes de niveau de service peuvent être introduites selon le type de clients ou le motif de l'appel. Les algorithmes intègrent alors des niveaux de priorité ou de qualité de service objectifs. Au-delà de la distribution d'appels, certaines fonctionnalités peuvent être

apportées aux clients telles que l'annonce du temps d'attente (Aguir, Chauvet et al. 2003) ou la possibilité de laisser un message et d'être rappelé.

Horizon	Problématique	Décisions
<p><u>Stratégique</u></p> <p>Long terme (années)</p> <p>Investissement et choix structurants</p>	<p>Mettre en place un centre d'appels opérationnel en adéquation avec les objectifs de l'entreprise</p>	<p>Infrastructures et technologies</p> <ul style="list-style-type: none"> • Localisation du site (ou des sites) • Acquisition du matériel informatique et télécom • Choix des fournisseurs <p>Organisation</p> <ul style="list-style-type: none"> • Positionnement en centre de coût ou centre de profit • Les indicateurs de performance • Externalisation ou internalisation • Flux et canaux traités • Profil de compétence des conseillers • Politique RH
<p><u>Tactique</u></p> <p>Moyen terme (semaines/ mois)</p> <p>Gestion des ressources humaines</p>	<p>Garantir l'adéquation entre charge et capacité tout en respectant les principales contraintes de temps de travail</p>	<p>Dimensionnement</p> <ul style="list-style-type: none"> • Ressources humaines (plan de recrutement) <p>Planification</p> <ul style="list-style-type: none"> • Positionnement des jours de repos et des jours de congé des conseillers • Détermination des emplois du temps des conseillers <p>Ajustement de la capacité de service</p> <ul style="list-style-type: none"> • Recrutements complémentaires (intérim, CDD) • Sollicitation de prestataires
<p><u>Opérationnel et temps réel</u></p> <p>(heures/ secondes)</p> <p>Pilotage temps réel de l'activité</p>	<p>Distribuer les appels et gérer les imprévus</p>	<ul style="list-style-type: none"> • Réactivité du management (procédures en cas de surcharge d'appels, d'incidents techniques, etc.) • Décisions de routage effectuées par les algorithmes opérant au sein des équipements du centre d'appels

Tableau I-1 : Les horizons de décision dans un centre d'appels

I.1.3 Focus sur l'horizon tactique : la chaîne de décision en gestion des ressources humaines

La masse salariale est le premier poste budgétaire d'un centre d'appels. Elle représente entre 60% et 80% des coûts (Coffre 2005). De ce fait, la gestion des ressources humaines est le principal levier pour garantir l'adéquation entre la charge d'appels prévisionnelle et la capacité de service mise en place.

Compte tenu de l'importance des coûts salariaux, le fonctionnement usuel dans les centres d'appels est de surdimensionner toutes les ressources matérielles (ressources informatiques, réseaux, etc.) et de piloter au plus juste la masse salariale en fonction de la demande. Ceci conduit à des problématiques d'organisation du temps de travail qui doivent intégrer différents types contraintes : l'obligation de disponibilité vis à vis des clients, les contraintes légales ou contractuelles, les desiderata des conseillers, etc.

Ces différentes problématiques induisent un grand nombre de décisions qui doivent être prises à différents horizons temporels. Chaque décision fige un peu plus le temps de travail des conseillers (durée du contrat, congés, jours de repos, horaires, etc.). Le jalonnement de ces décisions dans le temps varie selon les domaines d'activités, les contraintes légales ou les habitudes de l'entreprise.

Afin de donner une vision plus concrète de l'enchaînement et de l'imbrication de ces différentes décisions, nous allons nous attarder sur l'organisation illustrée dans la Figure I-3. Il ne s'agit que d'un exemple d'organisation du temps de travail dans un centre d'appels. Nous verrons dans le prochain chapitre que d'autres modes de fonctionnements peuvent être envisagés.

Le processus de gestion du temps de travail décrit dans la Figure I-3 est décomposé en plusieurs étapes. Chaque étape fournit des informations d'entrée à celle qui lui succède.

- **Dimensionnement** : il s'agit de déterminer le besoin en ressources humaines et effectuer les recrutements.
- **Planification journalière individuelle** : il s'agit de déterminer les jours d'absence et de présence des conseillers de clientèle. Dans notre exemple ceci se fait par un processus itératif où les conseillers expriment des choix (affectation calendriers) en fonction des contraintes sur la capacité de service présente.
- **Planification horaire agrégée** : Une fois les jours de présence déterminés, reste à planifier les horaires des conseillers. Dans notre exemple, ceci commence par une première étape de planification horaire agrégée, où l'entreprise détermine le

nombre de conseillers par vacation en fonction de la charge d'appels. Une vacation est un horaire type qui n'est pas encore affecté à un conseiller.

- **Emploi du temps :** il s'agit ici d'affecter les vacations aux conseillers. Généralement, l'objectif est de respecter au mieux les desiderata des conseillers. Les processus pour y arriver varient d'une entreprise à l'autre. Certaines entreprises laissent au superviseur le soin de répartir de la manière la plus équitable les horaires. On peut également utiliser des outils informatiques assez complexes permettant de recueillir les ordres de préférence des conseillers et de générer les emplois du temps maximisant leur satisfaction.

Chaque étape de ce processus constitue un problème à part entière qui suscite sa propre littérature scientifique. Dans cette thèse, nous allons nous intéresser particulièrement à un modèle, le *shift scheduling*, qui comme nous le verrons dans ce qui suit, est à la fois le centre de grands enjeux économiques (du fait qu'il représente l'essentiel des coûts d'un centre d'appels) et présente par ailleurs un intérêt scientifique.

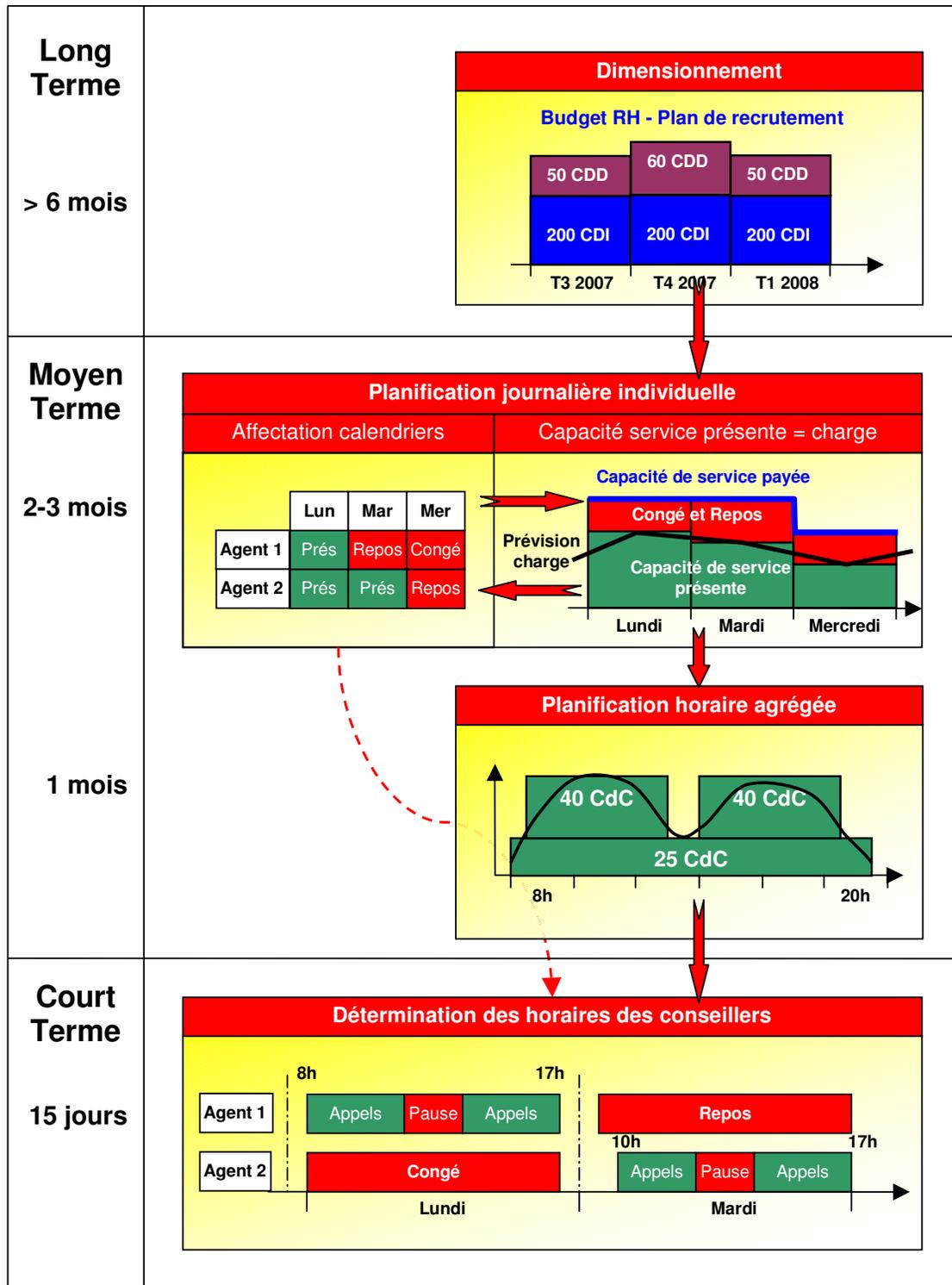


Figure I-3: Les problématiques de gestion des ressources humaines (*workforce management*). Exemple d'organisation dans un centre d'appels

I.2 Enjeux du *shift-scheduling*, un problèmes central dans l'optimisation des centres d'appels

Dans le monde des centres d'appels le *shift scheduling* est un problème central à la fois d'un point de vue économique et d'un point de vue scientifique.

I.2.1 Enjeu industriel : un problème au cœur de l'adéquation entre charge et capacité

Le *shift scheduling* assure l'adéquation entre la charge d'appels et la capacité de service au niveau horaire. En effet, il a pour objectif de garantir que, tout au long de la journée, la capacité de production est suffisante pour répondre à la demande.

La problématique d'adéquation entre charge et capacité se retrouve dans quasiment tous les secteurs d'activités industrielles ou de service. Deux éléments distinguent les centres d'appels des autres secteurs de l'industrie et rendent cette adéquation plus difficile et plus coûteuse à réaliser. D'une part, la charge d'appels est très variable au cours de la journée. Elle est difficilement compatible avec les contraintes de présence des effectifs. D'autre part, la réponse à un appel ne peut pas être différée. Elle doit se faire en temps réel. Ceci restreint considérablement la marge de flexibilité pour adapter la capacité à la charge d'appels.

I.2.1.1 Une charge d'appels à forte variabilité

Une personne appelle son centre de relation client lorsqu'elle est disponible. Généralement hors de son temps de travail. A une échelle collective, ceci conduit le centre d'appels à recevoir des pics d'appels au moment des pauses déjeuners ou à la fin des journées de travail (voir Figure I-4).

Au delà de la nécessaire adaptation à une charge variable, ceci impose au centre d'appels de mettre en place une organisation du temps de travail en total décalage avec les usages habituels. Pour répondre aux appels il faut donc prévoir le plus d'effectifs dans les plages habituellement « inactives ».

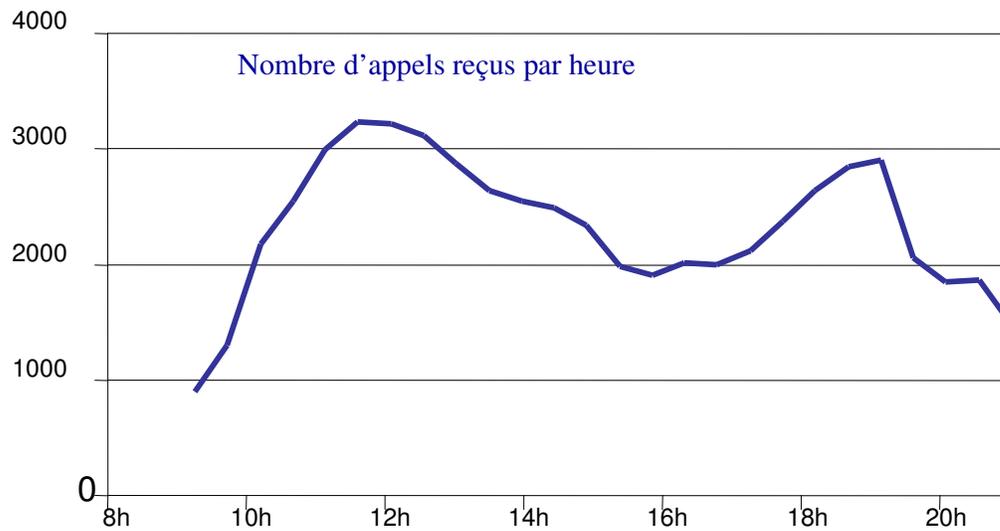


Figure I-4 : La charge d'appels n'est pas monotone au cours de la journée. Elle est très variable et présente plusieurs pics (source : charge d'appels d'un lundi typique pour un opérateur français de téléphonie mobile)

I.2.1.2 Une obligation de réponse en temps réel

Lorsqu'un appel arrive, il doit être traité quasi immédiatement. En effet, au delà de quelques minutes d'attente le client s'impatiente et raccroche. L'appel est à ce moment perdu. Cette contrainte d'immédiateté entre l'expression de la demande et sa satisfaction restreint considérablement la flexibilité en gestion des ressources humaines. Par opposition au monde des centres d'appels, une flexibilité existe dans l'industrie par le fait qu'il est possible de différer le moment de la production du moment de l'expression de la demande. Lorsque la production se fait avant l'expression de la demande, on parle de production sur stock. Lorsque la production se fait après la demande, on parle de production sur commande (Baglin, Bruel et al. 2001). Ces deux moyens de flexibilité n'existent pas dans un centre d'appels : on ne peut évidemment pas répondre à un client avant qu'il ait appelé et on ne peut pas non plus attendre au delà de quelques minutes pour lui répondre.

L'impact de cette contrainte d'immédiateté sur l'adéquation entre charge et capacité est illustré dans la Figure I-5. On voit que dans le cas où on a la possibilité de produire sur stock et sur commande, toute la demande peut être satisfaite si elle ne dépasse pas la capacité de production sur la journée. Dans le cas d'un centre d'appels, on constate qu'une partie de la demande n'est pas satisfaite et que la capacité n'est pas exploitée pleinement.

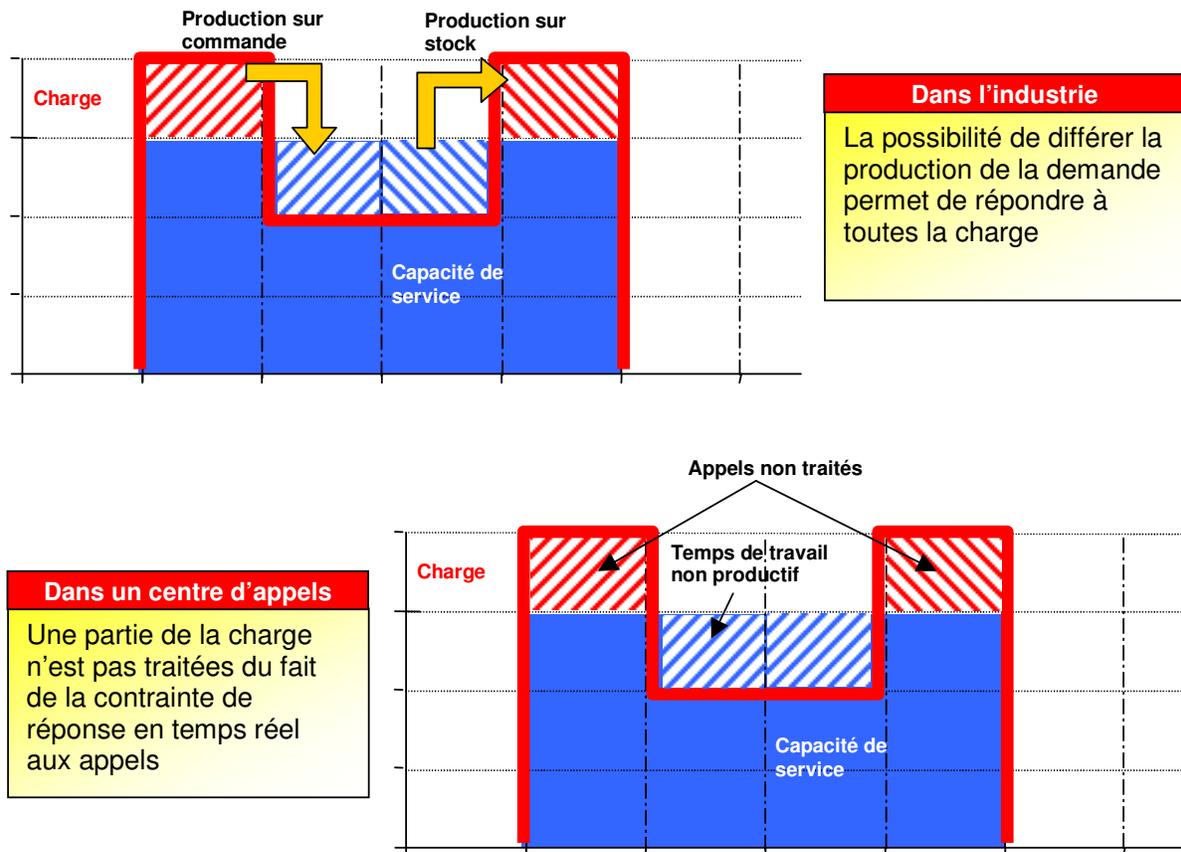


Figure I-5 : Impact de la contrainte de traitement en temps réel de la demande sur l'adéquation entre charge et capacité dans un centre d'appels par opposition à l'industrie

En réalité, il existe d'autres moyens de flexibilité dans les centres d'appels. Nous pouvons citer l'utilisation de flux stockable (mail et courrier, par exemple) qui se prêtent à un traitement différé ou alors les actions de communication afin d'inciter les clients à appeler lors des périodes d'activité creuse du centre d'appels. N'oublions pas également d'évoquer le moyen de flexibilité le plus courant dans le monde des centres d'appels qui consiste simplement à ne pas prendre en charge les appels et inviter les clients à rappeler ultérieurement.

I.2.2 Enjeu scientifique : un modèle d'optimisation de plusieurs problèmes de gestion des ressources humaines

Nous avons vu que le shift-scheduling abordait des enjeux importants d'un point de vue économique pour un centre d'appels. Au delà de son intérêt industriel, le shift-scheduling est quasiment incontournable pour qui s'intéresse aux modèles mathématiques d'optimisation de la gestion des ressources humaines d'un centre d'appels. En effet, la modélisation mathématique en shift-scheduling permet de résoudre plusieurs problèmes de gestion des ressources humaines. Notamment :

- Le problème de dimensionnement où il s'agit de déterminer le nombre d'agents qu'il est nécessaire de recruter pour répondre à la demande. La première modélisation du shift-scheduling a d'ailleurs été proposée sur un problème de dimensionnement d'agent de péage dans (Dantzig 1954).
- Le problème de planification permettant de déterminer le nombre de conseillers qui doivent être présents à chaque période pour maximiser la qualité de service offerte au client.
- Le problème de détermination des emplois du temps : nous montrons dans cette thèse que le modèle de planification que nous proposons permet d'aborder les problématiques d'emplois du temps individuels des conseillers.

Pour conclure, nous pouvons dire que le problème de *shift-scheduling* est très important à la fois d'un point de vue économique et scientifique. Il sera, en quelque sorte, le fil conducteur de cette thèse. L'objectif étant de répondre à une problématique industrielle tout en proposant des éléments nouveaux d'un point de vue de recherche scientifique.

I.3 Problématiques : les axes de contribution de la thèse

A la lumière d'une première analyse de l'état de l'art et d'une expérience opérationnelle dans le domaine des centres d'appels, nous avons identifié trois axes de recherche sur lesquels cette thèse contribuerait. Dans ce qui suit nous déclinons ces trois axes de recherche qui correspondent au trois chapitres de la suite du document.

I.3.1 L'incomplétude de l'information, un principe au fondement de l'organisation des décisions

Nous avons montré dans la section I.1.2 que les décisions de gestion de ressources humaines dans un centre d'appels étaient hiérarchisées en plusieurs niveaux de décision. Ces différents niveaux donnent lieu à plusieurs problématiques distinctes (voir Figure I-3) : dimensionnement, planification, emplois du temps individuels, etc.

Dans la littérature, cette décomposition est souvent prise comme donnée de base. La pertinence ou les fondements de cette décomposition sont très peu discutés. A notre connaissance, cette question n'a été abordée dans la littérature que d'un point de vue algorithmique et technique. C'est à dire comparer le gain en temps de calcul qu'apporte la décomposition d'un problème par rapport à la perte d'optimalité que ceci engendre. De notre point de vue, l'argument technique (optimalité vs temps de calcul) n'est pas suffisant pour justifier les décompositions. D'autres raisons, en particulier organisationnelles, doivent entrer en ligne de compte. En effet, ce principe de décomposition des décisions n'est pas spécifique aux centres d'appels. On le retrouve dans tout secteur d'activité y compris dans des processus ne nécessitant pas d'outils informatiques pour lesquels aucun problème de temps de calcul ne se pose. Nous pouvons citer par exemple, les approches hiérarchique et de consolidation utilisées pour l'affectation des budgets dans pratiquement toutes les entreprises et administrations. Il y a donc d'autres raisons sous-jacentes, notamment organisationnelles, derrière ce principe de décomposition au delà des simples raisons techniques.

Le chapitre 2 tente d'éclaircir ces raisons organisationnelles. Le caractère incomplet et incertain de l'information est, de notre point de vue, l'élément qui conduit à retarder et décomposer les décisions de gestion des ressources humaines. Nous montrons à partir d'exemples illustratifs comment la décomposition de décisions permet de pallier les inconvénients d'une information incomplète.

I.3.2 Proposer un paradigme et un modèle unificateur du shift-scheduling

La première modélisation en PLNE du problème de shift-scheduling a été proposée par Dantzig (Dantzig 1954). Depuis, de très nombreux travaux ont été réalisés pour adapter le modèle et permettre la résolution de problèmes de plus en plus complexes (nous présentons un état de l'art des modèles dans le chapitre 3). Toutefois, chaque contribution propose un modèle spécifique en fonction des hypothèses qu'elle prend en compte : avec ou sans pauses déjeuner, possibilité d'avoir des pauses hors déjeuner, durée des pauses fixes ou variables, nombre de profils de conseillers, un ou plusieurs flux traités, contraintes de durée maximum de travail, etc.

Dans cette thèse, nous présentons un concept que nous appelons **le paradigme de chaînes d'activités**. Ce paradigme permet :

- de modéliser en un programme mathématique le shift-scheduling quelles que soient ses spécificités (pauses, contrainte de durée maximum ou minimum de travail) pour peu que les contraintes de temps de travail respectent un certain nombre d'hypothèses que nous spécifierons ;
- de s'affranchir du formalisme mathématique pour traduire tout problème de shift scheduling. C'est la raison pour laquelle on parle de paradigme au delà du modèle mathématique qui en découle. Ceci est particulièrement utile dans la conception d'interfaces d'outils de planification et simplifie le dialogue avec les utilisateurs opérationnels.

Le chapitre 3 présente le paradigme de chaîne d'activité. Le modèle mathématique qui en découle et son application à un problème de dimensionnement seront également présentés dans ce même chapitre.

I.3.3 Prendre en compte la qualité de service de manière explicite

Dans un centre d'appels, la qualité de service mesure la facilité d'accès au service. Les indicateurs couramment utilisés sont le temps moyen d'attente, la proportion d'appels répondus en moins de 20 secondes ou encore la proportion d'appels répondus.

Dans le problème de *shift scheduling*, la question de la qualité de service est généralement traitée en amont. L'approche classique utilisée initialement par Dantzig se fait en deux temps. Dans un premier temps, on évalue sur chaque période de la journée le nombre de

conseillers nécessaires pour atteindre un objectif de qualité de service donné. Le *shift scheduling* peut alors être résolu dans un second temps en prenant directement comme contrainte ce nombre de conseillers présents par période. On parle de contraintes de couverture.

Le *shift scheduling* étant un problème combinatoire, traiter la qualité de service en amont, permet de découpler l'aspect combinatoire du modèle et la non-linéarité de la qualité de service. Ceci donne in fine un problème plus simple à résoudre.

Cette approche basée sur des contraintes de couverture "dures" (i.e. devant être respectées à tout prix) a prévalu dans la quasi-totalité des contributions traitant du *shift scheduling* depuis Dantzig. Même si elle permet de simplifier la résolution, elle n'est pas sans présenter quelques inconvénients. En effet, dans un centre d'appels, il n'est pas impératif de garantir un seuil de qualité de service constant durant la journée. On pourra se permettre de compenser les "mauvaises" périodes par de meilleures périodes de la journée. Dans ce cas, il serait plus pertinent de considérer, par exemple, une qualité de service globale sur la journée. Cette approche serait plus réaliste mais rendrait le problème plus difficile à résoudre. En effet, ceci impliquerait d'introduire dans la combinatoire du *shift scheduling* une évaluation non linéaire de la qualité de service.

Dans le chapitre 4 nous montrons comment, grâce au paradigme de chaîne d'activités, nous pouvons résoudre le problème de *shift scheduling* avec un objectif de maximisation de la qualité de service non linéaire.

Chapitre II

ORGANISATION DES DÉCISIONS DE GESTION DES RESSOURCES HUMAINES

Ce chapitre questionne l'organisation des décisions dans un centre d'appels. Généralement, dans les travaux visant à optimiser la gestion des ressources humaines, les problèmes sont présentés dans leur formulation mathématique sans justification ou remise en cause des processus de décision sous-jacents. Ce chapitre tente d'expliquer et justifier les décompositions possibles des décisions dans un centre d'appels. Elles sont, à notre sens, d'abord liées à des raisons d'organisation, de disponibilité de l'information et de contraintes industrielles.

II.1 Méthodologie d'organisation des décisions de gestion des ressources humaines dans les centres d'appels

Nous avons montré dans le chapitre précédent un exemple de décomposition des décisions de gestion des ressources humaines. Dans la littérature, ce type de décomposition est pris comme hypothèse de base. Nous tentons de montrer, dans ce qui suit, que les décompositions des décisions dans une organisation suivent une logique. Le principe sous-jacent à notre sens est celui de **l'incomplétude de l'information**. En effet, au moment de la prise d'une quelconque décision, l'information sur le futur n'est jamais totalement connue. Cette information est incomplète pour deux raisons principales :

- **L'incertitude des prévisions** : tout processus de décision pour préparer l'avenir doit se baser d'une manière ou d'une autre sur des prévisions. Dans un centre d'appels par exemple, ces prévisions concernent la charge d'appels, le nombre de courriers à traiter, le taux d'absentéisme, etc. Plus l'horizon de prévision est éloigné, plus les erreurs sont importantes. L'impact sur la qualité de la décision peut alors être très significatifs.
- **La prise de décision par plusieurs acteurs** : pour des raisons de satisfaction des collaborateurs notamment, il est très courant que les décisions de gestion du temps de travail (planification journalière, emploi du temps) se prennent conjointement par l'entreprise et les collaborateurs. Par exemple, la détermination des emplois du temps peut se faire itérativement entre le responsable hiérarchique et les conseillers de manière à concilier les objectifs de l'entreprise de réponse aux clients et les objectifs des conseillers d'obtenir des emplois du temps compatibles avec leurs contraintes personnelles.

Afin d'explicitier le lien que nous voyons entre l'incomplétude de l'information et l'organisation des décisions de gestion des ressources humaines nous commencerons par prendre un cas théorique déterministe. Nous tenterons de montrer que dans une situation où l'information est incomplète, il est judicieux de décomposer et retarder les décisions.

Par ailleurs, nous nous focaliserons dans un premier temps sur les décisions de gestion du temps de travail (dimensionnement et planification), nous montrerons par la suite l'impact de l'introduction des décisions d'ajustement de capacité (intérim, prestataires, etc.) .

II.1.1 Cas déterministe avec une décision centralisée

Afin d'illustrer les raisons de la décomposition, commençons par prendre un cas théorique d'un centre d'appels où trois décisions de gestion du temps de travail doivent être prises :

- Le dimensionnement des ressources humaines
- Le calendrier des conseillers (planification journalière) : c'est-à-dire la détermination des jours de présence et d'absence (congé et repos)
- Les emplois du temps des conseillers (planification horaire)

Nous supposons de plus que toutes ces décisions sont prises par l'entreprise à elle seule. Il n'y a pas, pour le moment, de processus impliquant le conseiller dans l'organisation de son temps de travail.

Considérons, enfin, que l'environnement est totalement déterministe. Ceci signifie que l'avenir est connu avec certitude et que l'on dispose bien à l'avance de la volumétrie précise d'appels à chaque instant.

Long Terme	<div style="border: 1px solid black; background-color: #4CAF50; color: white; padding: 5px; text-align: center;">Dimensionnement Calendrier (jours de repos et congé) Emplois du temps</div>
Moyen Terme	
Court Terme	

Figure II-1: Les décisions de gestion des ressources humaines dans un environnement déterministe

Dans ce cas, toutes les décisions de gestion du temps de travail (recrutement, calendrier, emplois du temps) peuvent être prises très longtemps à l'avance (horizon long terme). De plus, il ne devient plus nécessaire de recourir à des ajustements de capacité de service étant donné qu'en l'absence d'incertitude, les décisions optimales prises à long terme le restent à court terme. Le schéma de fonctionnement pertinent serait alors celui illustré dans la Figure II-1. **Dans un environnement déterministe, mono-décideur et totalement prévisible il n'y aucun intérêt à décomposer ou retarder une quelconque décision de gestion des ressources humaines.**

II.1.2 Le principe de décomposition - impact de l'incertitude

Considérons à présent un cas plus réaliste où l'environnement est incertain. L'information sur la demande est d'autant plus imprécise que l'horizon de temps est éloigné. Ainsi, dans ce cas de figure, l'entreprise aura tout intérêt à retarder au maximum les décisions de gestion des ressources humaines. L'objectif étant de prendre ces décisions lorsque la connaissance de la demande est la plus précise (voir Figure II-2)

Long Terme	
Moyen Terme	
Court Terme	<div style="border: 1px solid black; background-color: #4CAF50; color: white; padding: 5px; text-align: center;"> Dimensionnement Calendrier (jours de repos et congé) Emplois du temps </div>

Figure II-2: Impact de l'incertitude - retarder les décisions au maximum

Ainsi, **dans un contexte incertain, il est de l'intérêt de l'entreprise de retarder toutes les décisions.** Toutefois, les contraintes organisationnelles, opérationnelles ou légales font que tout ne peut pas être décidé au dernier moment. Le recrutement, par exemple, est une décision lourde. Elle nécessite un processus préalable de prospection, d'identification et de sélection de candidats et se prépare plusieurs mois à l'avance. Elle ne peut être effectuée que sur du moyen ou long terme. En ce qui concerne les congés, le droit du travail impose de les diffuser aux collaborateurs au moins un mois à l'avance. Certaines conventions collectives peuvent aller bien au-delà. Retarder toutes les décisions comme illustré dans la Figure II-2 n'est donc pas réalisable.

A partir de ces deux exemples, il apparaît assez clairement que deux préoccupations opposées conditionneront l'organisation et la hiérarchisation des décisions de gestion des ressources humaines au sein d'une entreprise :

- **Une nécessité d'anticiper** afin de maximiser la marge de manœuvre de l'entreprise pour garantir une adéquation entre charge et capacité. Cette marge de manœuvre est d'autant plus réduite que l'on se rapproche du court terme (à court terme par exemple, on ne peut plus bénéficier de la même marge de manœuvre offerte par des recrutements).

- **Une nécessité de prendre les décisions au plus tard** afin de tirer parti du niveau d'information le plus précis possible. Ce niveau d'information augmente à mesure que l'on se rapproche du court terme.

Ces deux tendances sont illustrées par deux courbes qui évoluent en sens opposés dans la Figure II-3. Les deux exemples ci-dessus (Figure II-1 et Figure II-2) correspondent aux cas extrêmes de chacune de ces tendances.

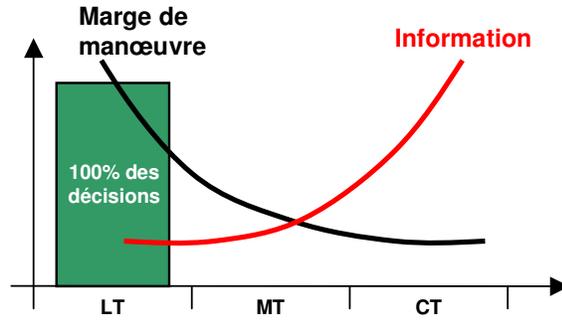
Le premier exemple où toutes les décisions sont positionnées à long terme offre une marge de manœuvre maximale à l'entreprise mais avec un niveau d'information au minimum. Ce cas de figure ne se rencontre quasiment jamais en pratique. En effet, fonder la prise de toutes les décisions sur des informations très peu fiables conduit nécessairement à des résultats médiocres en terme d'adéquation charge capacité.

Retarder, en revanche, toutes les décisions au court terme permettrait de maximiser le niveau d'information. C'est le cas illustré par le second exemple mais qui n'est pas réalisable.

La solution de compromis entre ces deux cas extrêmes est de tirer profit du fait que toutes les décisions n'ont pas les mêmes contraintes temporelles. Il est, de ce fait, envisageable de ne retarder qu'une partie des décisions pour bénéficier d'une information plus fiable tout en restant dans un cadre faisable. C'est le principe de décomposition des décisions.

Dans ce qui suit nous présentons quelques exemples de décomposition en appliquant ce principe.

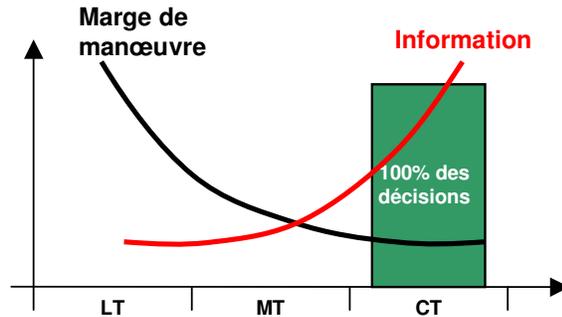
LT	100% décisions
MT	
CT	



a) Maximisation de la marge de manœuvre

Sous-optimale : toutes les décisions sont prises lorsque l'information est minimum

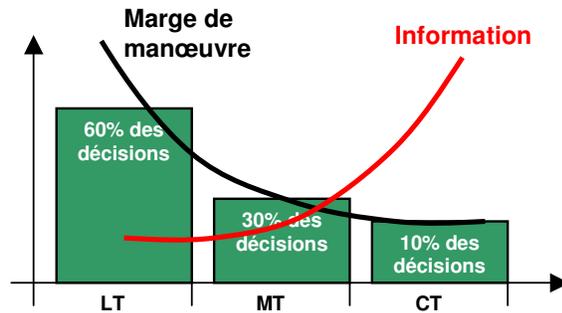
LT	
MT	
CT	100% décisions



b) Maximisation du niveau d'information

Non-réalisable : La marge de manœuvre à CT n'est pas suffisante pour mettre en œuvre toutes les décisions

LT	Phase 1
MT	Phase 2
CT	Phase 3



c) Maximiser le niveau d'information compte tenu de la marge de manœuvre possible

La décomposition permet de retarder une partie des décisions pour obtenir un meilleur niveau d'information

Figure II-3: Gestion du compromis entre niveau d'information et marge de manœuvre dans gestion des ressources humaines

II.1.2.1 Décomposition dimensionnement-planification

Reprenons notre exemple initial et supposons que la seule contrainte temporelle en terme de marge de manœuvre est celle du dimensionnement qui doit se faire à long terme. On considère également que l'entreprise n'a toujours pas recours à des actions d'ajustement de capacité de service (intérim, CDD, etc.).

Dans ce cas, l'application du principe de décomposition est représentée dans la Figure II-4 où le dimensionnement est effectué à long terme. Toutes les autres décisions sont retardées soit au court terme s'il n'y a pas d'autres contraintes temporelles, ou au moyen terme (1 à 3 mois) lorsque l'on a des contraintes légales qui imposent la diffusion des calendriers et des emplois du temps au moins un mois à l'avance.

Long Terme	Dimensionnement	Long Terme	Dimensionnement
Moyen Terme		Moyen Terme	Planification (calendrier + emplois de temps)
Court Terme	Planification (calendrier + emplois de temps)	Court Terme	

Contrainte de diffusion du calendrier aux collaborateurs à moyen terme

Figure II-4 : Impact de l'incertitude - retarder les décisions au maximum avec une contrainte temporelle sur le dimensionnement

II.1.2.2 Décomposition de la planification journalière et horaire

Nous avons vu précédemment que les contraintes légales de diffusion des calendriers plusieurs mois à l'avance conduisaient à avancer la planification du court terme au moyen terme. A ce niveau, une nouvelle décomposition peut s'avérer pertinente. En effet, une nouvelle possibilité serait de maintenir la création des calendriers ou planification journalière à moyen terme afin de respecter la contrainte légale et de retarder l'établissement des emplois du temps à court terme bénéficiant ainsi de l'amélioration de la prévision. Cette décomposition supplémentaire est illustrée dans la Figure II-5.

Long Terme	Dimensionnement
Moyen Terme	Calendrier (jours de repos et congé)
Court Terme	Emplois du temps

Figure II-5 : Impact de l'incertitude - retarder les emplois du temps au court terme

Nous avons montré ici quelques exemples de décompositions des décisions qui pouvaient être induits par l'incertitude sur les prévisions. Il ne s'agit que d'exemples illustratifs et d'autres types de décompositions peuvent être utilisés. Par exemple, il est envisageable de ne figer qu'une proportion des emplois du temps à moyen terme. Le reste est planifié au plus tard lorsque les prévisions sont plus fiables. Les conseillers connaissent ainsi l'essentiel de leurs emplois du temps plusieurs semaines à l'avance et l'entreprise se réserve une marge de flexibilité qui peut être réajustée si besoin.

II.1.3 Le principe de décomposition – impact du partage de la décision

Jusque là nous avons supposé que l'entreprise était la seule à décider de sa planification pour optimiser l'adéquation entre charge et capacité. En plus de cette préoccupation de service au client, l'entreprise doit également intégrer la satisfaction de ses collaborateurs. Rappelons que dans les centres d'appels, les entreprises doivent octroyer toute son importance aux conditions de travail. C'est un levier essentiel de motivation des conseillers, main d'œuvre généralement peu qualifié et faiblement rémunérée. Il est donc naturel et judicieux d'impliquer le conseiller dans la gestion de son temps de travail.

On passe alors d'un problème mono-décideur où l'entreprise décide des calendriers et emplois du temps de ses conseillers à un problème multi-décideurs faisant intervenir le conseiller dans l'organisation de son temps de travail. Ce partage de la décision entre l'entreprise et le conseiller conduit à un second type de décomposition des décisions. Dans la Figure II-6 nous proposons deux exemples illustrant ce principe :

- **Décomposition de la détermination des calendriers :** l'entreprise fixe pour chaque jour le nombre de conseillers présents nécessaires pour satisfaire la demande. Les conseillers positionnent leurs choix en fonction de cette contrainte.

- **Décomposition de la détermination des emplois du temps** : l'entreprise propose des horaires type que l'on appelle vacations et détermine le nombre de conseillers nécessaires par vacation. Les conseillers choisissent parmi ces possibilités les horaires qui leur correspondent le mieux.

Si certains jours de présences ou d'absences ou certaines vacations sont plus demandés que d'autres, l'entreprise peut imposer des règles ou des critères d'attribution ou de priorité pour garantir l'équité de traitement entre les collaborateurs et éviter que certains soient plus favorisés que d'autres au fil du temps. Il s'agira par exemple de fixer des quotas mensuels ou trimestriels sur le nombre de « long week-end » pris par conseiller ou encore de répartir équitablement les horaires les moins appréciés (début de la journée de travail très tôt ou fin très tardive)

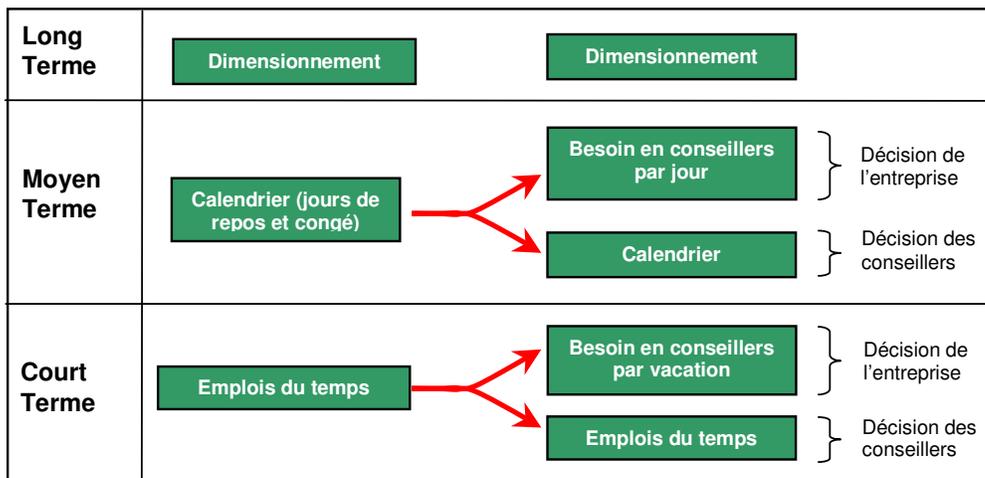


Figure II-6 : décomposer les décisions dans un fonctionnement de multi-décideurs

L'implication de plusieurs décideurs (en l'occurrence l'entreprise et les conseillers) dans la gestion du temps de travail est un choix organisationnel. C'est donc un facteur interne à l'entreprise qui conduit à décomposer le processus de décision. Par opposition, l'incertitude des prévisions est un facteur externe à l'entreprise lié au caractère aléatoire de son environnement.

II.1.4 Limites de la décomposition des décisions

Nous avons vu que la décomposition des décisions était le moyen utilisé au sein d'une organisation pour limiter les conséquences d'une information incomplète. Dans certains cas cependant, cette solution n'est pas judicieuse.

En effet, dans un cas totalement déterministe et mono-décideur, la solution d'un problème pris dans sa globalité est toujours meilleure ou égale que la résolution séparée des sous-problèmes qui le composent. La décomposition d'une décision n'a de sens qu'en la confrontant à la perte d'optimalité qu'elle induit. Si cette dégradation de la qualité de la solution est supérieure au gain apporté par des sous-décisions retardées, alors il n'est pas pertinent de décomposer.

La traduction mathématique de ce principe est illustrée dans la Figure II-7.

Moyen Terme		$\begin{cases} (X_a^*, Y_a^*) = \operatorname{argmin}_{X,Y} \hat{f}(X,Y) \\ \text{S.C } \hat{S}(X,Y) \end{cases}$	$\begin{cases} X_b^* = \operatorname{argmin}_{X,Y} \hat{f}(X,\bullet) \\ \text{S.C } \hat{S}(X,\bullet) \end{cases}$
Court Terme	$\begin{cases} (X^*, Y^*) = \operatorname{argmin}_{X,Y} f(X,Y) \\ \text{S.C } S(X,Y) \end{cases}$		$\begin{cases} Y_b^* = \operatorname{argmin}_{X,Y} f(X_b^*, Y) \\ \text{S.C } S(X_b^*, Y) \end{cases}$
	Cas idéal non faisable	a) Pas de décomposition X et Y déterminés à long terme	b) Décomposition en 2 sous-décisions X déterminé à long terme et Y à court terme

Figure II-7 : Illustration du compromis entre résoudre un problème global ou le décomposer en deux sous-décisions

Considérons le problème Pb_0 définit comme suit :

$$\begin{cases} \operatorname{Min}_{X,Y} f(X,Y) \\ \text{S.C } S(X,Y) \end{cases}$$

L'objectif est de minimiser $f(X, Y)$ où X et Y sont deux vecteurs de variables de décision.

$S(X, Y)$ correspond à l'ensemble des contraintes qui doivent être respectées

La solution optimale de ce problème est notée (X^*, Y^*)

Si aucune contrainte n'imposait le moment où les décisions doivent être prises, alors la détermination de X et Y doit se faire au plus tard (c'est à dire à court terme) pour disposer de l'information la plus précise.

Supposons à présent que, pour des raisons opérationnelles ou organisationnelles, X ne peut être fixé qu'à moyen terme au plus tard. Le problème Pb_0 n'est alors plus applicable. Deux alternatives se présentent :

a) Avancer au moyen terme les décisions sur X et Y

Notons (X_a^*, Y_a^*) la solution optimale du nouveau problème. Il se formule de la manière suivante :

$$\begin{cases} \text{Min}_{X, Y} \hat{f}(X, Y) \\ \text{S.C } \hat{S}(X, Y) \end{cases}$$

Les fonction f et S ont été remplacées par leur estimateurs \hat{f} et \hat{S} car, dans ce cas, les décisions reposent sur des estimations du futur (i.e. des prévisions).

b) Décomposer la décision entre court terme et moyen terme

Dans ce cas, X est déterminé à moyen terme et Y à court terme. Ceci nous conduit à deux sous-problèmes Pb_{MT} et Pb_{CT} dont la solution optimale est (X_b^*, Y_b^*) . Ces deux sous-problèmes peuvent se formuler de la manière suivante :

$$\begin{array}{ll} a) Pb_{MT} \text{ à moyen terme} & b) Pb_{CT} \text{ à court terme} \\ \begin{cases} \text{Min}_{X, Y} \hat{f}(X, \bullet) \\ \text{S.C } \hat{S}(X, \bullet) \end{cases} & \begin{cases} \text{Min}_{X, Y} f(X_b^*, Y) \\ \text{S.C } S(X_b^*, Y) \end{cases} \end{array}$$

Cette décomposition n'est pertinente que si le cas b) est plus favorable que le cas a). C'est à dire si :

$$f(X_b^*, Y_b^*) \leq f(X_a^*, Y_a^*)$$

II.1.5 Introduction des décisions d'ajustement de capacité

Jusque là, nous avons considéré qu'une fois les décisions de planification prises, il n'y avait plus de possibilité de les remettre en cause. Etant donné que certaines décisions peuvent être prises longtemps à l'avance - par exemple, la phase de dimensionnement dans l'exemple précédent - l'entreprise peut être amenée à les remettre en cause plus tard. Ainsi, des recrutements en CDI décidés 6 mois à l'avance, n'ont de pertinence qu'en fonction de la qualité des prévisions dont on dispose. Avec le temps les prévisions s'affinent et le dimensionnement établi peut s'avérer inadapté pour garantir une bonne adéquation entre charge et capacité. Dans ce cas, l'entreprise peut être amenée à augmenter la capacité de service via des actions ponctuelles telles que le recrutement d'intérimaires, de CDD ou la sollicitation de prestataires externes. A l'instar des décisions de gestion du temps de travail (dimensionnement et planification), les leviers d'ajustement de la capacité de service ont leurs propres contraintes temporelles comme indiqué dans le Tableau II-1.

Horizon	Décisions de gestion du temps de travail	Décisions d'ajustement de la capacité de service
> 6 mois	Budget RH et plan de recrutement (nombre de CDI, CDD, etc.) Recrutement à long terme (CDI)	
De 1 à 3 mois	Positionnement des jours de repos et de congé	Recrutement CDD Recrutement d'intérimaires Sollicitation d'un prestataire
15 jours	Emplois du temps	Heures supplémentaires
Quelques heures à quelques jours		Réaffectation d'activités Sollicitation de conseillers d'autres services en renfort

Tableau II-1 : Les contraintes temporelles des décisions de gestion des ressources humaines

Lors de l'introduction d'une décision d'ajustement de capacité, deux cas de figure peuvent se présenter : soit cette décision n'interfère pas avec les autres décisions. Il ne s'agit alors que d'une étape supplémentaire. Soit cette décision est liée à d'autres et son introduction implique une modification du processus de décision.

Comme précédemment, prenons un exemple fictif afin d'illustrer ces deux cas. Considérons un centre d'appels dont le processus de gestion du temps de travail est constitué de deux étapes : le dimensionnement à long terme et la construction des calendriers et emplois du temps à court terme (voir Figure II-8 (a)). Considérons, de plus, que l'entreprise souhaite avoir recours à des intérimaires ponctuellement. Ces intérimaires doivent être recrutés au moins un mois à l'avance (moyen terme). Les deux options envisageables sont les suivantes :

- Soit le nombre d'intérimaire peut être fixé indépendamment des autres décisions et on introduit une étape supplémentaire dans le processus sans autre altération (Figure II-8(b))
- Soit l'entreprise estime que la détermination du nombre d'intérimaire ne peut être dissociée du problème d'établissement des calendriers et emplois du temps. Ce serait le cas, par exemple, si des contraintes contractuelles ou organisationnelle imposaient de figer les emplois du temps des intérimaires dès la proposition de recrutement. Ceci imposerait d'avancer la réalisation des calendriers et des emplois du temps au même horizon que le recrutement des intérimaires, c'est à dire au moyen terme au lieu du court terme (Figure II-8(c)) .

L'entreprise est donc, à nouveau, face à un choix entre deux manières différentes de jalonner les décisions. Le choix dépendra de la qualité de la solution que peut donner chaque option et des contraintes organisationnelles ou contractuelles que doit respecter l'entreprise.

Remarque : notons que l'on peut introduire un niveau de complexité supplémentaire en considérant que les décisions d'ajustement de capacité (qu'elles soient associées ou pas à des décisions de gestion du temps de travail) peuvent elles mêmes faire l'objet de nouvelles décompositions en appliquant les principes décrits en II.1.2, II.1.3 et II.1.4

	(a)	(b)	(c)
Long Terme	Dimensionnement	Dimensionnement	Dimensionnement
Moyen Terme		Recrutement intérim.	Calendrier Emplois du temps + recrutement intérim.
Court Terme	Calendrier Emplois du temps	Calendrier Emplois du temps	

Figure II-8 : Impacts sur le processus de décision de l'introduction de décisions d'ajustement de capacité

II.1.6 Synthèse

Nous avons montré dans ce qui précède que l'élément sous jacent à la décomposition est d'abord organisationnel avant toute considération technique de temps de calcul ou d'exhaustivité des éléments pris en compte dans les modèles. En effet, afin de pallier l'impact de l'incomplétude de l'information (environnement incertain et décision multi-acteur) nous avons identifié quelques lignes directrices pour organiser les décisions :

- Retarder les décisions au plus tard afin de bénéficier de l'information la plus précise
- Décomposer les décisions en sous décisions si l'optimalité n'est pas pénalisée et décaler au plus tard les sous-décisions obtenues
- Intégrer les leviers d'ajustement de capacité et adapter le processus de décision en conséquence

La Figure II-9 donne une illustration de cette démarche de constitution du processus de décision. Il ne s'agit, bien entendu, que d'un exemple. En réalité, il n'existe pas d'organisation idéale et unique des décisions de gestion des ressources humaines car elle dépend de plusieurs facteurs :

- La flexibilité de l'organisation permettant de décomposer plus ou moins facilement les décisions.
- L'incertitude et la variabilité de l'environnement.
- Les contraintes réglementaires et organisationnelles de l'entreprise.

	Environnement Déterministe	Environnement Incertain				
		Pas de décision MT à prendre	Décision à 3 mois : contrainte légale de diffusion des congés			
			Décomposition impossible sans dégradation importante de l'optimalité	Flexibilité suffisante pour décomposer sans perte d'optimalité		
				Pas de décision à 1 mois à prendre	Décision à 1 mois : intérim, prestataire	
					Décideur unique : entreprise	Décision conseiller et entreprise
LT > 6 mois	Dimensionnement Calendrier et Emplois du temps	Dimensionnement	Dimensionnement	Dimensionnement	Dimensionnement	Dimensionnement
MT 3 mois			Calendrier et Emplois du temps	Calendrier	Calendrier	Besoin en conseillers présents Calendrier
MT 1 mois					Emplois du temps + intérim, presta	Besoin en conseillers par vacation + intérim, presta
CT 15 jours		Calendrier et Emplois du temps		Emplois du temps		Emplois du temps

Figure II-9 : Les problématiques de gestion des ressources humaines. Exemple illustrant le processus de décomposition des décisions

II.2 Etat de l'art des problèmes de gestion des ressources humaines dans les services

Les problématiques de gestion des ressources humaines dans les centres d'appels et plus généralement dans les services font l'objet d'une littérature scientifique très riche. Plusieurs classifications et typologies de problèmes ont été proposées ainsi qu'un grand nombre de décompositions. Ces classifications se font en général sur la base des spécificités et contraintes prises en compte dans le problème ou sur la nature des méthodes de résolution utilisées (Tien et Kamiyama 1982; Günes 1999; Ernst, Jiang et al. 2004; Canon 2005). A notre connaissance, il n'y a pas, dans le domaine des services, de contribution semblable à celle que nous avons proposée plus haut décrivant la logique sous-jacente à l'organisation des décisions de gestion des ressources humaines.

Cette partie présentera quelques éléments de l'état de l'art relatif aux problématiques de gestion des ressources humaines. Il ne s'agira pas de dresser une bibliographie exhaustive mais de proposer un aperçu des types de modèles et méthodes qui permettent de répondre à ces problématiques. L'objectif étant in fine de mettre en perspective notre démarche de décomposition des décisions par rapport aux travaux et approches classiquement utilisées dans la littérature.

Parmi les travaux effectués dans le domaine de la gestion des ressources humaines, les problématiques de planification sont, de loin, les plus abordées. En ce qui concerne les problématiques de dimensionnement c'est à dire l'aide à la décision pour le recrutement à long terme, des travaux de référence peuvent être trouvés dans (Holt C., Modigliani F. et al. 1960; Grinold et Marshall 1977; Charnes, Cooper et al. 1978). Ils proposent des méthodes de mathématiques appliquées pour optimiser des problématiques de recrutement et de plan de carrière permettant de minimiser les coûts salariaux. D'autres travaux ont intégré, au delà de l'aspect coût, des préoccupations de maintien et de transmission des compétences sur le long terme (Bordoloi et Matsuo 2001).

En ce qui concerne la planification, plusieurs revues de littératures existent. Une classification intéressante est celle proposée par (Partouche 1998) qui distingue trois types de problème :

- Planification avec tâches courtes ou sécable (à courbe de charge)
- Planification de personnel mobile (problème de construction de service, bus)
- Planification de personnel mobile avec découché (avion, construction de rotation)

Nous nous focaliserons sur la 1^{ère} catégorie car c'est celle qui s'applique aux centres d'appels. Sur ce type de problématiques, nous pouvons citer les problèmes et modèles les plus couramment abordés dans la littérature. Pour des éléments plus détaillés se référer à (Partouche 1998) et (Baker 1976) :

- *Shift-scheduling* (problème de construction de vacation): planification au sein d'une journée à partir d'une courbe de charge. Les contraintes liant les journées entre elles (la contrainte de 35 heures hebdomadaires, par exemples) ou les successions de vacations (par exemple, imposer 2 jours de repos successifs dans la semaine) ne sont pas prises en compte.
- *Days-off scheduling* (problème de planification de jours de repos): concerne les activités de service sur 6 ou 7 jours par semaines mais dont les salariés ne travaillent que 5 ou 6 jours. Le problème du *days-off scheduling* permet de positionner le ou les jours de repos hebdomadaires pour chaque personne.
- *Tour scheduling* (problème de planification de tours) : mix entre les deux problèmes précédents. Il s'agit de construire les vacations et les jours de repos en même temps sur un horizon d'une semaine.
- *Problème global de construction de grille* : défini pour la première fois dans (Partouche 1998). Il consiste à déterminer toutes les décisions de planification journalière et horaire. En prenant en entrée une courbe de charge, l'objectif du problème est de déterminer : l'affectation individuelle des jours de repos, les vacations à construire, l'affectation du nombre de personnes aux vacations et les emplois du temps individuels.

Un élément important à noter dans cette classification est que les problèmes sont constitués en fonction des éléments pris en compte et des données d'entrée du problème. C'est une classification orientée résolution et non pas décision. En réalité, un même problème mathématique peut être utilisé pour répondre à différentes décisions qui ont lieu à différents horizons temporels. Ceci est illustré dans la Figure II-1 où nous reprenons un des exemples de décomposition des décisions décrit plus haut. Pour chaque décision, nous indiquons un exemple de modèle qui peut être utilisé pour répondre à la question posée. Le *shift-scheduling* par exemple n'est pas exclusivement destiné au problème de planification. Il peut être utilisé à la fois pour une problématique de recrutement à long terme et pour l'établissement des emplois du temps. Bien entendu, il ne serait pas pertinent d'utiliser la même finesse dans les deux cas. Par exemple, on introduira moins de contraintes de temps de travail pour le dimensionnement que pour la détermination des emplois du temps. Ainsi, pour une décision donnée, plusieurs modèles plus ou moins complexes existent dans la littérature. De la même manière, un même modèle peut être utilisé pour répondre à différents types de décisions.

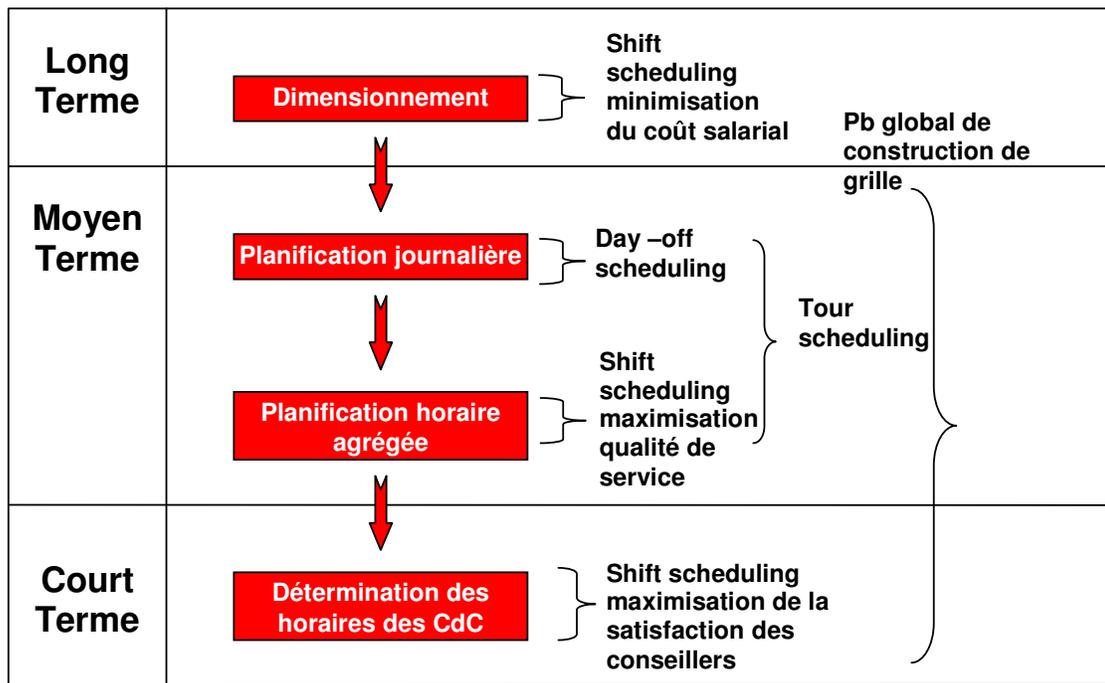


Figure II-10: Positionnement des problèmes de planification à courbe de charge par rapport à la hiérarchisation des décisions de workforce management

Avant d'aborder un problème de gestion des ressources humaines il est donc nécessaire de suivre une analyse en deux temps. Dans un premier temps, définir l'organisation et le jalonnement des décisions en suivant l'approche décrite en II.1.6. Dans un second temps, on cherchera à identifier les moyens les plus pertinents pour répondre à ces décisions. Le degré de complexité de ces solutions pourra aller du logiciel intégré issu des modèles mathématiques d'optimisation présentés dans la littérature à des méthodes beaucoup plus simples et opérationnelles : simple accord verbal ou écrit avec la hiérarchie, tableau d'affichage des horaires, fichiers Excel, etc.

Chapitre III

DIMENSIONNEMENT SOUS CONTRAINTE DE COUVERTURE

Ce chapitre développe les principes du paradigme de chaîne d'activités. C'est un paradigme que nous utilisons pour représenter, modéliser et résoudre le problème de *shift scheduling*. Afin d'en illustrer les principes, nous appliquons notre paradigme à la modélisation d'un problème de dimensionnement d'un centre d'appels.

Après une présentation du problème, nous proposons une revue bibliographique des modélisations du *shift scheduling*. Nous présentons, par la suite, les contours du paradigme de chaîne d'activités et le modèle qui en découle. Nous généralisons également ce modèle aux centres d'appels traitant plusieurs types de flux en plus des appels. Nous terminons enfin par une analyse de la structure et de la difficulté de notre problèmes et nous évaluons une borne d'une solution approchée.

III.1 Position du problème

Nous nous intéressons, dans ce chapitre, au problème de dimensionnement d'un centre d'appels. Comme illustré dans la Figure III-1, le dimensionnement est la première décision à prendre en matière de gestion des ressources humaines. Les décisions de planification (positionnement des congé et jours de repos et affectation des emplois du temps) sont prises ultérieurement.

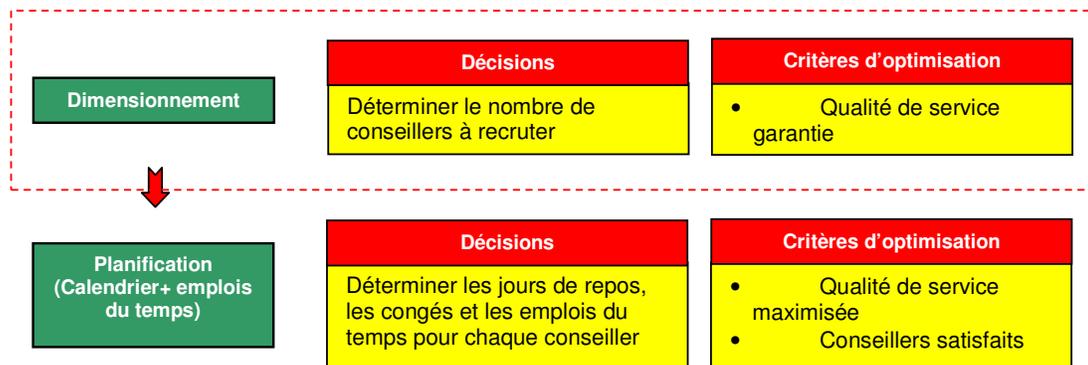


Figure III-1 : Processus de décision englobant le problème de dimensionnement

L'objectif du dimensionnement est de déterminer le nombre de conseillers à recruter en fonction de la qualité de service que l'entreprise souhaite garantir. Pour ce faire, plusieurs aspects sont à considérer :

- Les prévisions d'appels,
- les contraintes de temps de travail,
- les imprévus tels que l'absentéisme,
- la qualité de service à garantir, etc.

Chaque entreprise possède son propre mode de fonctionnement pour prendre en compte ces différents aspects. Dans ce qui suit, nous considérons que la détermination du nombre de recrutements se fait selon le processus décrit dans la Figure III-2. C'est le processus classiquement utilisé dans le secteur des centres d'appels.

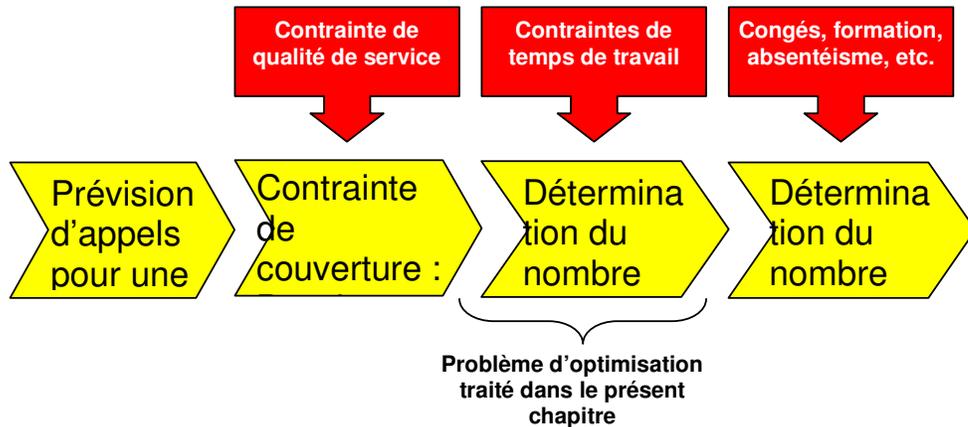


Figure III-2 : Processus sous-jacent à la construction de la décision de recrutement

Ce processus se décompose en plusieurs étapes :

a) Prévision d'appels sur une journée type

La première étape consiste à construire les prévisions d'appels. Le dimensionnement se faisant généralement sur un horizon temporel assez long, il n'est pas possible de disposer de prévisions très précises. Aussi, on prend habituellement comme référence une journée type de l'historique. Ceci permet d'estimer le nombre d'appels par période sur une journée représentative du fonctionnement habituel d'un centre d'appels.

b) Détermination des contraintes de couverture

Une fois les prévisions d'appels établies, le but de cette étape est d'évaluer pour chaque période, le nombre de conseillers nécessaires pour garantir l'objectif de qualité de service que s'est fixée l'entreprise. L'usage habituel consiste à imposer une contrainte de qualité de service identique pour chaque période de la journée. Cette contrainte est par la suite convertie en nombre de conseillers nécessaires. Ceci se fait généralement en ayant recours à des modèles de files d'attente (Baker 1976). La Figure III-3 illustre graphiquement ce type de méthodes pour deux critères de qualité de service : la proportion d'appels répondus et le temps moyen d'attente.

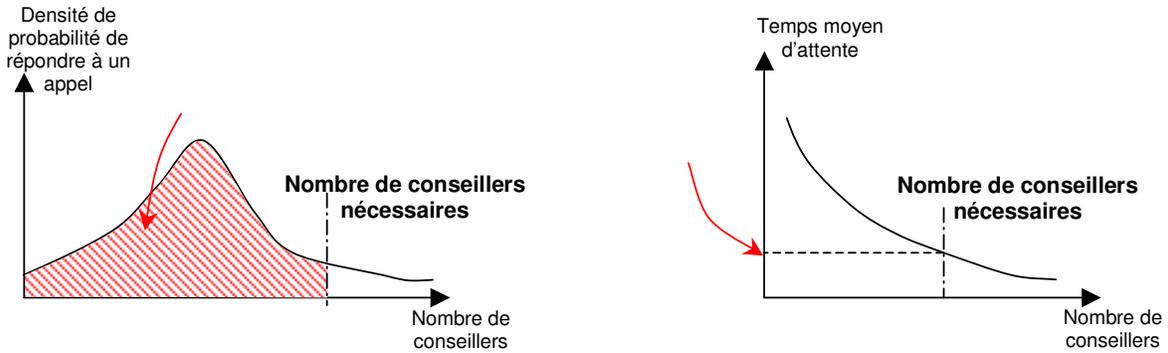


Figure III-3 : Conversion d'une contrainte de qualité de service en nombre de conseillers nécessaires

c) Détermination du nombre de conseillers présents

Il s'agit de la problématique qui est au cœur du présent chapitre. Cette étape consiste à déterminer le nombre de conseillers nécessaires sur la journée type identifiée de manière à minimiser le coût salarial tout en respectant les contraintes de qualité de service. Il est également nécessaire de s'assurer que la solution est applicable. Pour cela, un certain nombre de contraintes de temps de travail doivent également être respectées : temps de travail maximum, heure d'arrivée au plus tard, etc.

d) Détermination du nombre de conseillers à recruter

Une fois que le nombre de conseillers présents est évalué pour une journée type, il reste à déterminer le nombre réel de conseillers qui devront recruter. Généralement des règles empiriques sont utilisées par les entreprises. Une approche classique consiste à appliquer des coefficients correctifs afin de prendre en compte les éléments qui n'ont pas été intégrés dans le problème de l'étape précédente : congés, absentéisme, variations saisonnières du volume d'appels, etc. Nous ne nous attarderons pas sur les subtilités ou la pertinence de ces approches car elles varient considérablement en fonction de l'environnement, des contraintes et des pratiques de l'entreprise (Cleveland et Mayben 1997; ICMI 2000).

Dans ce qui suit, nous nous intéresserons au problème de détermination du nombre de conseiller présents sur une journée type de manière à minimiser le coût salarial (3^e étape du processus décrit dans la Figure III-3). C'est un problème largement abordé dans la littérature et connu sous le nom de problème de construction de vacation ou *shift scheduling*. Nous commencerons par présenter un état de l'art des modèles utilisés pour formuler et résoudre ce problème. Nous présenterons par la suite le **paradigme de chaîne d'activité** qui nous permet de formuler et de résoudre le problème de *shift scheduling* sous une forme très générique.

III.2 Etat de l'art des modélisations du problème de construction de vacations (*shift scheduling problem*)

III.2.1 Définition du problème de construction de vacations

Le problème de construction de vacations est un problème de planification de personnel. Il est connu dans la littérature scientifique internationale sous le nom de *shift scheduling problem* (Garey et Johnson 1979). Il s'agit, pour une journée donnée, de déterminer une affectation des employés à un ensemble de vacations autorisées afin, d'une part de couvrir une courbe de charge traduisant une demande et d'autre part, de minimiser le coût de travail.

III.2.1.1 Quelques définitions :

Vacation (shift) : c'est en quelque sorte la "brique" élémentaire de planification à laquelle sont affectés les employés. C'est un ensemble de périodes de planification consécutives. Chaque vacation est caractérisée par un temps de début, un temps de fin et éventuellement une ou plusieurs pauses pour lesquelles sont également définis des temps de début et de fin (voir Figure III-4).

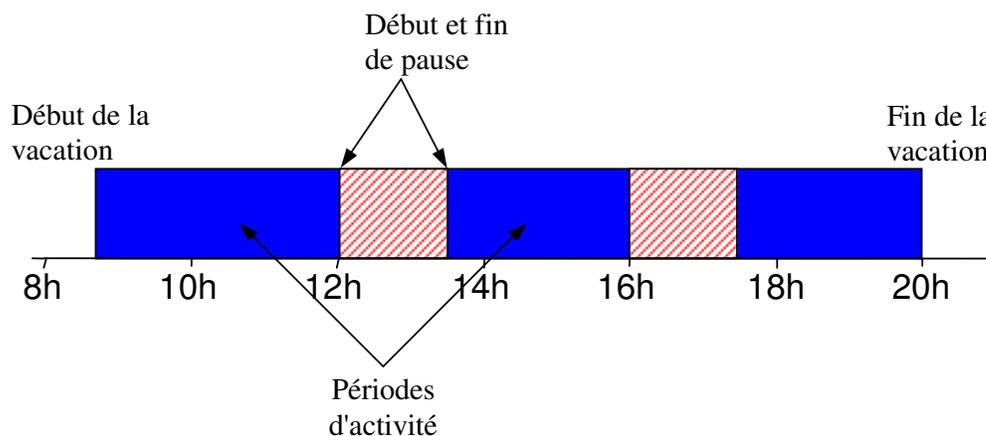


Figure III-4: Illustration de la définition d'une vacation

Période de planification : La journée est divisée en périodes de planification. C'est l'intervalle de temps élémentaire sur lequel on définit une charge à satisfaire. Les durées de travail ou de

pauses sont définies par un nombre entier de période de planification. Dans la pratique, cette période varie entre 15 et 60 minutes pour les centre d'appels (voire à des pas beaucoup plus fins de l'ordre de 5 mn dans des métiers tels que le transport aérien). Dans la majorité des études, les périodes sont de durées identiques. La plupart des approches peuvent être généralisées à des périodes de durées différentes au prix d'une complexification de la présentation. Notons que définir des périodes de planification conduit à **discrétiser le temps**.

Courbe de charge ou contrainte de couverture : Elle correspond au besoin en nombre d'employés nécessaires par période de planification. En général, elle est obtenue à partir d'estimations prévisionnelles de la demande. Par exemple, dans le cas d'un centre d'appels, il s'agit de prévision d'appels par période. En fonction du niveau de qualité de service que l'on désire obtenir (par exemple le temps moyen d'attente ou la proportion d'appels répondus sans attente) on peut évaluer le nombre nécessaire de conseillers par période. Comme nous l'avons déjà évoqué, ceci se fait généralement sur la base de modèles de files d'attente. Pour plus de détail sur la manière d'établir une courbe de charge à l'aide de modèles de files d'attente voir (Baker 1976; Koole et Mandelbaum 2002).

III.2.2 Les premiers modèles

III.2.2.1 Programme linéaire en nombres entiers (Dantzig 1954)

Dantzig a proposé (Dantzig 1954) une formulation en problème de couverture (*set covering problem*). Son modèle a été appliqué à une problématique proposée par (Edie 1954) où il s'agissait de déterminer le nombre minimal d'employés pour assurer le fonctionnement d'un poste à péages. La formulation mathématique du problème est la suivante :

$$\text{Min} \sum_{j=1}^n c_j x_j$$

sous les contraintes

$$\sum_{j=1}^n a_{ij} x_j \geq D_i \quad i = 1 \text{ à } m$$

$$x_j \in \mathbb{N}$$

i est l'indice de la période de planification considérée (allant de 1 à m)

j est l'indice de la vacation considérée (allant de 1 à n)

x_j correspond au nombre d'employés affectés à la vacation j (variable de décision à optimiser)

D_i correspond au nombre d'employés nécessaires pour la période i (contrainte de couverture)

c_j correspond au coût de travail d'un employé affecté à la vacation j

$$a_{ij} = \begin{cases} 1 & \text{si la période } i \text{ correspond à une période de travail pour la vacation } j \\ 0 & \text{sinon} \end{cases}$$

Le modèle consiste à déterminer le nombre d'employés x_j de chaque vacation j de manière à minimiser le coût global généré tout en couvrant la charge de chaque période i .

Limites du modèle

Le modèle de Dantzig permet de prendre en compte tout type de vacation avec un nombre quelconque de pauses. Toutefois, associer une variable de décision à chaque vacation peut augmenter la taille du modèle de manière considérable. Ceci est accentué lorsque les vacations sont définies avec une certaine flexibilité sur les horaires de travail des conseillers. En effet, en règle générale et pour avoir une plus grande marge de manœuvre, on préfère spécifier des fenêtres de temps pour le début et la fin de travail plutôt que des heures fixes. Ceci est également le cas pour les débuts et fins de pause. Pour tenir compte de cette flexibilité dans le modèle de Dantzig, il est nécessaire d'explicitier chaque vacation possible par une variable de décision. Par exemple, si on a k_1 périodes possibles pour le début d'activité, k_2 pour la fin, k_3 pour le début de pause, on devra expliciter $k_1 * k_2 * k_3$ vacations différentes dans le modèle, on voit donc que la taille du modèle augmente très vite avec la flexibilité. Pour pallier ce problème, des modèles dits implicites ont été proposés, ils sont présentés dans la partie III.2.3.

III.2.2.2 Problème de flot (Segal 1974)

Segal s'est inspiré du modèle de Dantzig pour proposer une méthode originale de résolution du problème de construction de vacations. Son approche consiste à utiliser un modèle de flot pour résoudre le problème sans tenir compte des pauses. A partir de cette première solution, il fait intervenir une procédure heuristique pour réintégrer les pauses. Les vacations sont toutes définies de manière explicite dans ce modèle.

Le modèle de flot est illustré dans la Figure III-5. Les nœuds représentent les instants de début et de fin des périodes de planification. Nous avons deux types d'arcs, les *forward arcs* de i à $i+1$ qui représentent les périodes de planification allant de l'instant i à $i+1$ et les *backward arcs* entre m et l ($l < m$) qui représentent les vacations. En effet, en l'absence de pause, chaque vacation est définie uniquement par son début l et sa fin m . La charge à couvrir durant la période i à $i+1$ est la capacité minimale de l'arc $(i, i+1)$. Et le coût unitaire d'affectation d'un employé à la vacation commençant en l et finissant en m est C_{ml} , le coût de l'arc (m, l) . La solution du problème sans pauses revient à trouver le flot de coût minimal. Le nombre d'employés affectés à la vacation commençant en l et finissant en m sera le flot passant par l'arc (m, l) . De plus, étant données les propriétés d'un problème de flot, si la charge à couvrir est entière alors il existe un flot optimal de valeurs entières (en nombre de conseillers).

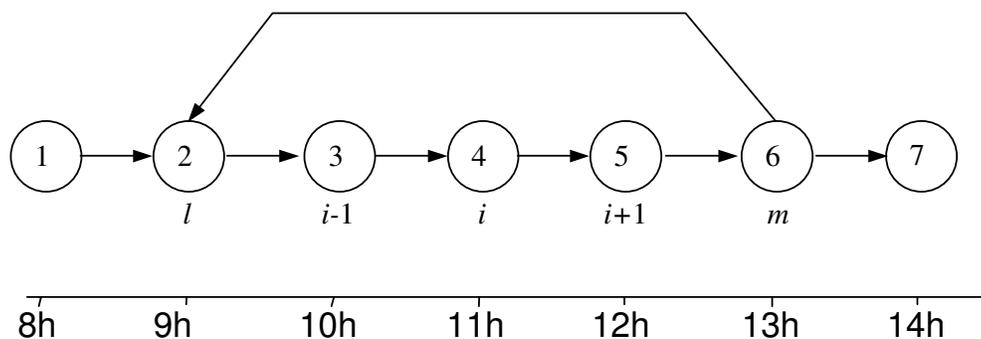


Figure III-5: Modèle de flot de Segal

Limites du modèle

L'approche de Segal est très intéressante dans la mesure où il met en évidence que le problème de construction de vacations sans pause et sans flexibilité se ramène à un problème de flot qui peut ainsi être résolu en un temps polynomial. Toutefois, l'approche qu'il propose pour réintégrer les pauses est basée sur une méthode heuristique. Cette approche, ne permet donc pas, dans le cas général, d'atteindre l'optimum.

III.2.3 Les modèles implicites

III.2.3.1 Le concept de modélisation implicite

Comme nous l'avons mentionné plus haut, le modèle de Dantzig présente l'inconvénient d'augmenter en taille à mesure que la flexibilité sur les vacations augmente (fenêtre de temps, durée maximale et minimale). Pour pallier ce problème, des modèles dits implicites ont été proposés dans la littérature. L'idée consiste à ne plus expliciter ces flexibilités dans les vacations en associant une variable de décision à chaque vacation possible mais plutôt à utiliser d'autres variables de décision dont le nombre ne dépend pas de la flexibilité. Dans ce qui suit, nous présenterons les contributions les plus significatives en terme de modélisation implicite.

III.2.3.2 Début, fin et durée de vacations implicites (Moondra 1976)

Moondra s'est intéressé au problème de construction de vacation des employés d'une banque. Il considère deux catégories d'employés, des employés temps plein qui prennent

une pause au cours de la journée et des employés mi-temps qui n'ont pas de pause mais travaillent moins longtemps. Le modèle qu'il propose est un programme linéaire en nombres entiers à l'instar de Dantzig, mais il s'en distingue par des variables de décision qui ne sont pas associées aux vacances. Il introduit, à la place, des variables de décision qui correspondent au nombre d'employés qui commencent et terminent leur vacation à chaque période de planification. Ainsi, le nombre de variables ne dépend plus de la flexibilité apportée aux vacances.

L'intérêt dans la contribution de Moondra, en termes d'approche implicite, réside dans sa modélisation de la flexibilité de l'organisation du temps de travail des employés à mi-temps. En effet, dans le problème qu'il considère, ces derniers peuvent travailler entre 4 et 7 heures et ont la possibilité de débiter à n'importe quelle heure de la journée. Si l'on considère un horizon de planification de 9 périodes d'une heure, ceci correspond à 6 possibilités pour la période de début et 4 pour la durée de travail. En utilisant l'approche de Dantzig, il serait nécessaire de distinguer 18 vacances c'est à dire autant de variables de décision. Dans le modèle de Moondra, il suffit de 13 variables de décision auxquelles il faut ajouter 9 contraintes pour représenter les mêmes possibilités.

L'approche implicite de Moondra ne prend pas en compte le cas où l'on a une flexibilité sur le positionnement de la pause. Il faudra attendre une quinzaine d'années plus tard pour que Bechtold & Jacobs apportent une réponse à cette question.

III.2.3.3 Une seule pause implicite (Bechtold et Jacobs 1990)

Dans leur modèle, Bechtold & Jacobs distinguent deux types de variables de décision : des variables pour représenter les vacances sans pause et des variables pour modéliser les pauses. Les vacances sans pauses sont modélisées de manière explicite, les pauses par contre sont définies implicitement. Elles représentent le nombre d'employés qui commencent leur pause à chaque période de planification indépendamment de la vacation à laquelle ils sont affectés.

Des contraintes dites *backward* et *forward* sont utilisées pour mettre en cohérence ces pauses avec les vacances auxquelles elles sont associées. (Bechtold et Jacobs 1996) ont démontré l'équivalence entre leur modèle implicite et la formulation de Dantzig si on suppose qu'il n'existe pas de chevauchement extraordinaire. Le chevauchement extraordinaire (*extraordinary overlap*) est défini comme l'existence d'au moins une fenêtre de pause totalement incluse dans une autre (voir Figure III-6). Cette hypothèse est nécessaire car si elle n'est pas vérifiée alors le modèle de Bechtold et Jacobs affectera en premier les pauses aux vacances ayant la plus petite fenêtre de pause avant de les affecter aux autres vacances.

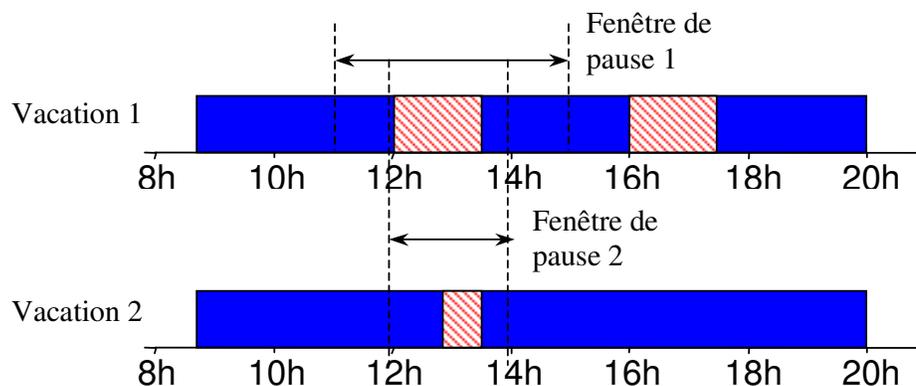


Figure III-6: Illustration du chevauchement extraordinaire

Bechtold et Jacobs ont montré expérimentalement que le fait de modéliser les pauses de manière implicite permettait de résoudre le problème plus rapidement qu'avec la formulation de Dantzig. Toutefois, le modèle présente encore quelques limites : contrairement au modèle de Moondra, les vacations sont définies de manière explicite. De plus, le modèle ne permet qu'une seule pause par vacation et les fenêtres de pause n'admettent pas de chevauchement extraordinaire. Le premier point sera résolu dans (Thompson 1995) et les deux derniers dans (Aykin 1996).

III.2.3.4 Modèle doublement implicite (Thompson 1995)

Le modèle de Moondra permet de modéliser de manière implicite la flexibilité sur le début, la fin et la durée des vacations. Le modèle de Bechtold & Jacobs considère quant à lui de manière implicite le positionnement d'une seule pause. (Thompson 1995) a intégré ces deux approches pour développer un modèle doublement implicite. Il introduit la notion de type de vacation que nous appelons dans ce document profil. Un type de vacation ou profil consiste en un ensemble de vacations ayant le même coût, la même durée de pause et les mêmes contraintes de temps de travail. A chaque profil et à chaque période de planification, il associe trois variables de décision : le nombre d'employés qui commencent leur journée de travail, le nombre d'employés qui la terminent et le nombre d'employés qui prennent leur pause. Les positionnements possibles de la pause sont déterminés en fonction de fenêtre de temps mais aussi en fonction des durées (minimales et maximales) permises pour la période d'activité qui la précède et celle qui lui succède.

Thompson teste expérimentalement sa formulation et la compare avec celle de Bechtold et Jacobs sur 588 problèmes. En moyenne sa formulation se révèle plus rapide à résoudre et permet d'aborder des problèmes de plus grande taille.

III.2.3.5 Plusieurs pauses implicites (Aykin 1996)

(Aykin 1996) prend en compte dans son modèle la possibilité d'avoir plusieurs pauses. Il conserve la modélisation explicite de Dantzig des vacances mais représente les pauses de manière implicite. En plus de prendre en compte plusieurs pauses, il se distingue de Bechtold et Jacobs en utilisant des variables de pauses différentes pour chaque vacation. Ceci fait que le problème de chevauchement extraordinaire ne se pose plus. De plus, Aykin propose également une généralisation de son modèle au cas de travail cyclique sur 24 heures.

Le modèle d'Aykin possède plus de variables de décision mais moins de contraintes que celui de Bechtold et Jacobs. Dans (Aykin 2000) sont présentés les résultats de tests comparatifs entre les deux modèles sur 220 problèmes. En règle générale, le modèle d'Aykin permet de résoudre des problèmes de plus grande taille et plus rapidement.

III.2.4 Synthèse des modèles

Le tableau suivant synthétise les différences et les apports de chaque modèle :

Modèle	Vacations hors pause		Pauses			Remarques
	explicités	implicites	explicités	Une pause implicite	Plusieurs pauses implicites	
Dantzig, 1954	✓		✓			- Problème de <i>set covering</i> - Explosion de la taille du problème avec la flexibilité des vacances
Segal, 1974	✓		✓			- Problème de flot - Pauses traitées séparément par une procédure heuristique
Moondra, 1976		✓	✓			
Bechtold, 1990	✓			✓		Variables de pause communes pour toutes les vacances
Thompson, 1995		✓		✓		Modèle doublement implicite : Combinaison des modèles de Moondra et Bechtold & Jacobs
Aykin, 1996	✓			✓	✓	Variables de pause distinctes pour chaque vacation

Tableau III-1 : Synthèse des principaux modèles de PLNE implicites et explicites pour la résolution du problème de construction de vacances

III.3 Modélisation via le paradigme de chaîne d'activité

Comme nous l'avons vu dans la partie précédente, l'approche de Dantzig nécessite d'énumérer toutes les vacations possibles dans le modèle. Cet aspect qui est considéré comme sa principale limite présente néanmoins l'avantage d'être générique. En effet, la représentation explicite des vacations permet, par construction, de traduire tout type de contrainte de temps de travail. Les modèles implicites qui ont été développés par la suite ne permettent pas, en revanche, la même généralité. Les contributions que nous avons présentées en III.2 introduisent des limitations sur le modèle (ne serait-ce que pour des raisons « pédagogiques » de clarté). Ainsi, on peut avoir, par exemple, des restrictions sur le nombre de pauses considérées, sur leur durée ou encore le nombre de types de contrats pris en compte. Ceci signifie que tout problème de construction de vacations qui n'entre pas exactement dans le cadre défini dans le modèle proposé nécessite une formulation spécifique (même s'il s'agit d'une modification mineure du modèle initial).

Le modèle que nous proposons tente de conserver autant que possible le caractère générique du modèle de Dantzig dans la traduction des contraintes de temps de travail tout en gardant les avantages d'une formulation implicite (la taille du modèle qui ne dépend pas du degré de flexibilité sur les vacations). Il repose sur un paradigme que nous appellerons le **paradigme de chaîne d'activité**.

Le paradigme de chaîne d'activité est une représentation non mathématique qui permet de traduire les contraintes de temps de travail de manière générique. Nous montrons qu'il est possible de construire à partir de ce paradigme un modèle implicite pour la résolution du problème de construction de vacation. De plus, nous montrons dans III.5 que le paradigme de chaîne d'activité permet d'aller au-delà des modélisations classiques du problème de *shift scheduling* en prenant en compte la multi-compétence des conseillers et les flux dits stockables (Chauvet, Nait-Abdallah et al. 2007).

III.3.1 Le paradigme de chaîne d'activité

Le paradigme de chaîne d'activité est illustré dans la Figure III-7. Il s'agit de représenter les vacations par une succession d'activités. Les conseillers, peuvent, en effet, enchaîner plusieurs activités au cours d'une journée de travail. L'ordre selon lequel les activités s'enchaînent est donné. Dans la Figure III-7, la séquence de travail du conseiller suit l'enchaînement suivant :

- prise d'appels

- pause déjeuner
- traitement de courrier
- prise d'appels
- fin de la journée de travail

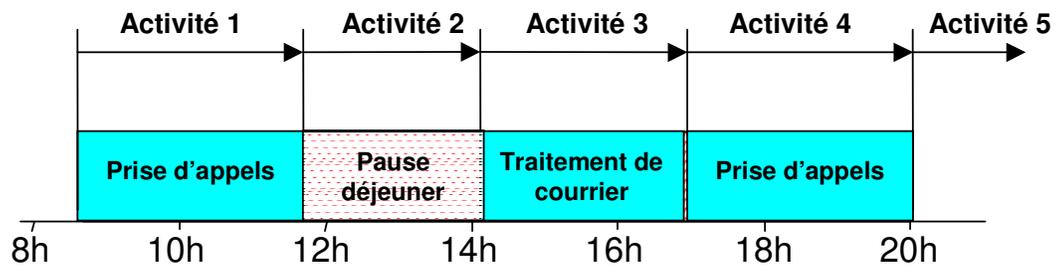


Figure III-7: Illustration du paradigme de chaîne d'activité

Les contraintes associées aux vacances ainsi qu'aux activités qui la composent (pauses incluses) sont de deux types (voir Figure III-8) :

- **Des contraintes de fenêtres de temps** définissant les dates au plus tôt et au plus tard de début ou de fin des vacances et des activités. Par exemple, un conseiller commence à travailler entre 8h et 10h, part en pause déjeuner entre 11h30 et 13h30 et commence à émettre des appels sortant entre 15h et 17h.
- **Des contraintes de durée maximale et minimale** permettant de limiter la durée d'une activité ou d'une suite d'activités. Par exemple, un conseiller ne peut pas répondre à des appels durant plus de 5 heures consécutives, la durée de sa pause déjeuner sera comprise entre 45 minutes au minimum et 2 heures au maximum et sa journée de travail ne dépassera pas 10 heures.

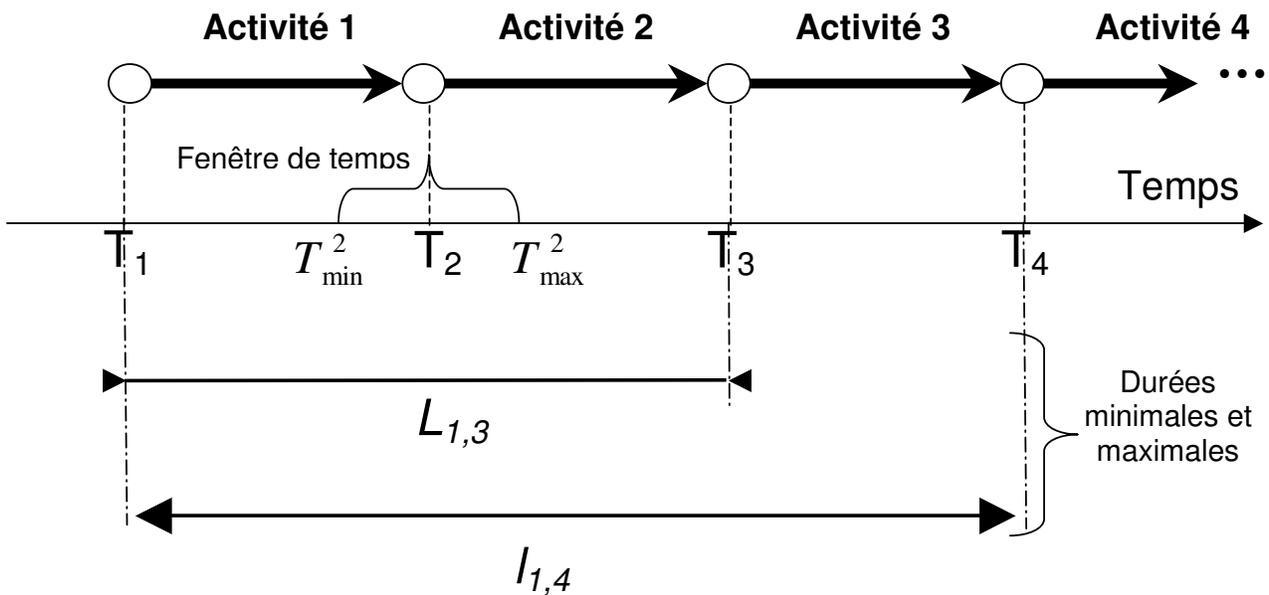


Figure III-8: Contraintes de temps de travail associées au paradigme de chaîne d'activité

$L_{1,3}$: délai maximal entre le début de l'activité 1 et le début de l'activité 3

$l_{1,4}$: délai minimal entre le début de l'activité 1 et le début de l'activité 4

T_{\min}^2 et T_{\max}^2 date au plus tôt et au plus tard pour le début de l'activité 2

On définit également la notion de profil qui regroupe les conseillers ayant les mêmes contraintes de temps de travail et le même coût salarial. Un profil est un ensemble de vacations constituées d'une même succession d'activités auxquelles est associé un ensemble de contraintes de fenêtres de temps et de durée. On pourra citer par exemple le cas d'un centre d'appels qui aurait deux profils. Un profil mi-temps correspondant à une durée maximale de présence de 4 heures sans pause. Et un profil temps-plein correspondant à une durée de présence comprise entre 6 heures au minimum et 9 heures au maximum avec une pause déjeuner d'une heure et deux pauses de repos d'une demi-heure (voir Figure III-9).

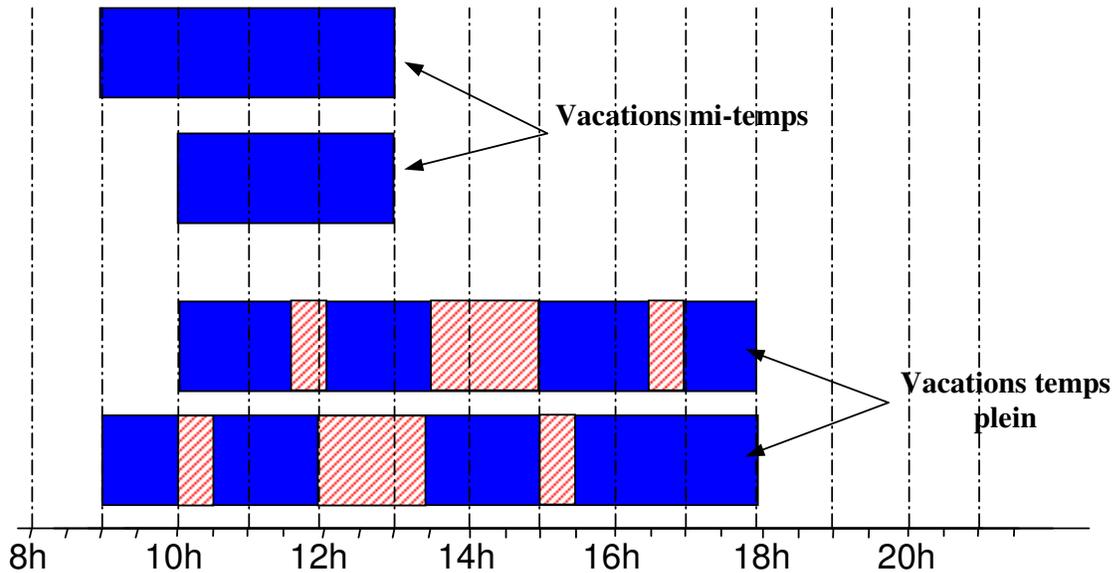


Figure III-9: Deux types de profils de vacances

III.3.2 Modélisation mathématique

Le modèle que nous proposons a été introduit dans le cas d'une seule pause dans (Chauvet, Chauvier et al. 2003) ainsi que dans (Rottembourg 2003), le paradigme de chaîne d'activité et l'intégration de plusieurs pauses ont été présentés dans (Chauvet, Chauvier et al. 2003; Chauvet 2005). Enfin la généralisation décrite dans la partie III.5 a été présentée dans (Chauvet, Nait-Abdallah et al. 2007).

III.3.2.1 Définitions et notations

Les définitions et notations associées au paradigme de chaîne d'activités sont illustrées dans la Figure III-10.

a) Un horizon de planification non cyclique

Nous supposons que la planification s'effectue sur un certain horizon. Sans perte de généralité, nous prendrons dans la suite un horizon d'une journée. La journée commence à la date T_0 et se termine à la date T_{end} . Ainsi, la première période de planification commence à T_0 et la dernière commence à $T_{end}-1$. Dans ce qui suit nous désignons par période t , la période qui commence à la date t . Le problème que nous abordons est non cyclique c'est à dire que la période d'activité est inférieure à 24h par jour.

b) Des conseillers qui effectuent une succession d'activités

Soit I l'ensemble des profils.

Pour chaque profil $i \in I$, on définit $A(i)$, l'ensemble des activités que peuvent effectuer les conseillers du profil i . Les éléments de cet ensemble sont ordonnés selon l'enchaînement des activités. Ainsi, si une activité $a \in A(i)$ alors $a+1$ est l'activité qui lui succède au sein du profil i .

Pour simplifier les formulations mathématiques ultérieures, nous introduisons la notion d'activité fictive a_{end} . Elle correspond à la fin de la journée de travail des conseillers du profil. Elle succède à la dernière activité « réelle » effectuée par les conseillers.

Pour l'instant nous considérons le cas où l'on a un seul type de flux à traiter : les appels. Une activité sera donc :

- soit une pause
- soit une activité de traitement d'appels
- soit l'activité fictive a_{end} correspondant à la fin de la journée de travail

Soit $A_{appels}(i)$ l'ensemble des activités de traitement d'appels du profil i .

Ainsi dans le cas où le seul flux à traiter est celui des appels, l'ensemble des pauses du profil i est :

$$A_{pause} = A(i) - A_{appels}(i) - \{a_{end}\}$$

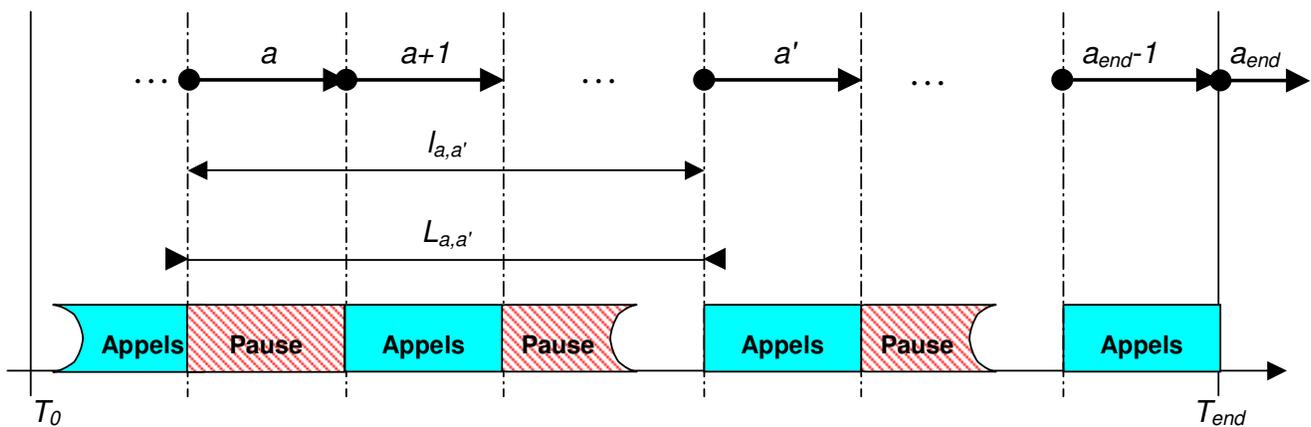


Figure III-10: Le paradigme de chaînes d'activités : définitions et notations

c) Les contraintes de temps de travail

Les contraintes de temps de travail sont définies comme suit

$l_{a,a'}(i)$ et $L_{a,a'}(i)$ correspondent aux durées minimale et maximale séparant les débuts des activités a et a' pour le profil i .

$[T_{\min}^a(i), T_{\max}^a(i)]$ correspond à la fenêtre de temps (dates au plus tôt et au plus tard) durant laquelle l'activité a doit commencer pour les conseillers du profil i .

III.3.3 Exemple d'application et limites du paradigme

III.3.3.1 Illustration du paradigme

Afin d'illustrer l'utilisation du paradigme de chaîne d'activités, étudions son application à l'exemple d'un centre d'appels avec un seul type de profil ayant les contraintes de temps de travail suivantes :

- Les vacances durent exactement 8 heures
- Les conseillers prennent une seule pause qui peut commencer entre 11h et 13h
- La pause dure entre 1h et 2h

Nous considérons de plus que la période de planification est égale à 1h (i.e. les activités sont planifiées par pas de 1h)

La transcription de ces contraintes dans le formalisme du paradigme de chaîne d'activités est illustrée dans la Figure III-11 :

- La journée de travail est décomposée en trois activités
 - Une première activité d'appels comprise entre l'étape 1 et 2
 - Une pause comprise entre l'étape 2 et 3
 - Une seconde activité d'appels comprise entre l'étape 3 et 4
- La contrainte de durée des vacances est formulée par une durée maximale et minimale entre les étapes 1 et 4 égales à 8h
- La fenêtre de temps (11h à 13h) pour le début de la pause est associée à l'étape 2

- Les contraintes sur la durée de la pause sont associées à la durée entre l'étape 2 et 3 : durée maximale de 2h et minimale de 1h

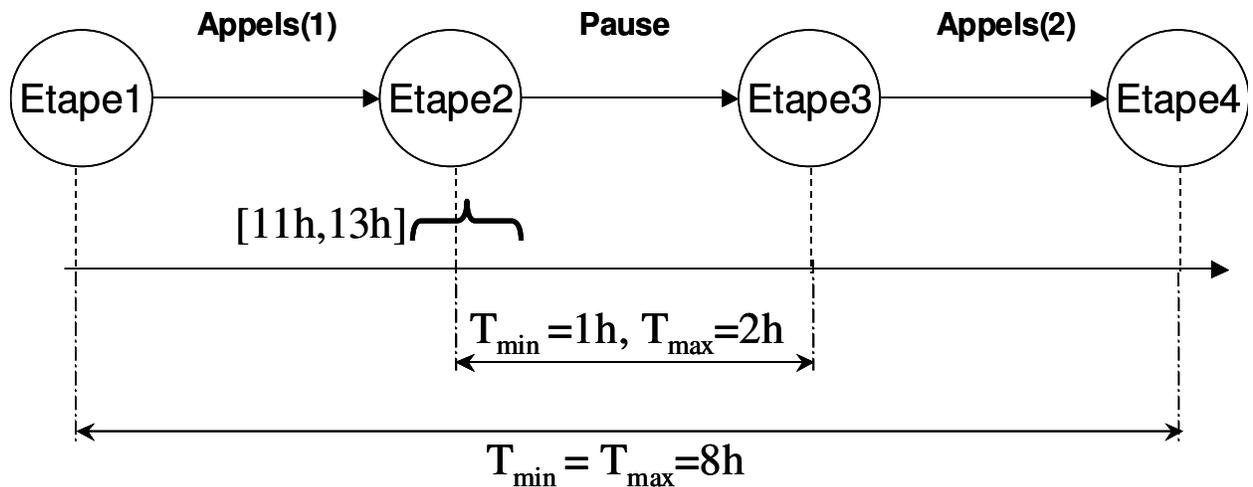


Figure III-11: Exemple d'application du paradigme de chaîne d'activités

III.3.3.2 Limites du paradigme de chaîne d'activités

Nous avons vu précédemment que les contraintes de temps de travail devaient être exprimées sur les étapes. Ainsi, il est possible d'imposer des contraintes sur la durée d'une activité ou la somme de plusieurs activités successives.

En revanche, le paradigme de chaîne d'activités n'offre pas de possibilité de formuler directement et de manière implicite des contraintes sur la somme des durées de plusieurs activités non successives. Pour y arriver, il est nécessaire de modéliser certaines parties du problème de manière explicite.

Pour illustrer cette limite, revenons à l'exemple précédent. Considérons à présent qu'il n'y a plus de contrainte sur la durée de la vacation. Nous la remplaçons par une durée des activités d'appels (Appels(1) + Appels(2)) égale à 7 heures. Dans ce cas, la durée de la pause ne peut plus être implicite. Pour prendre en compte cette contrainte, il est désormais nécessaire de créer 2 profils, chacun ayant une durée de pause fixe :

- Un premier profil avec une pause d'1 heure exactement; la contrainte de durée (8h) des activités d'appels revient à fixer la durée entre l'étape 1 et 4 à 8 heures

- Un second profil avec une pause de 2 heures exactement; la contrainte de durée (8h) des activités d'appels revient à fixer la durée entre l'étape 1 et 4 à 9 heures

III.3.4 Les variables de décision

Notons la variable de décision $X_t^a(i)$ qui correspond au nombre de conseillers affectés au profil i qui ont commencé l'activité a durant la période t ou avant. $X_t^{a+1}(i)$ correspond au nombre de conseillers qui ont terminé cette activité avant la date t . $X_t^a(i)$ est une variable de cumul croissante en fonction du temps.

Ces variables de cumul permettent **à elles seules** de déduire les autres informations liées à la planification des conseillers

Nombre de conseillers affectés à un profil

Soit $N(i)$ le nombre de conseillers qui seront affectés au profil i . Etant donné que les conseillers effectuent toutes les activités associées à un profil, on aura par définition l'égalité suivante :

$$N(i) = X_{T_{end}}^a(i) \quad \forall a \in A(i), \forall i \in I$$

Nombre de conseillers affectés à une activité

Soit $N_t^a(i)$ le nombre de conseillers d'un profil i dans l'activité a durant la période t . Il correspond à la différence entre le nombre de conseillers qui ont commencé et ceux qui ont terminé cette activité avant la période t . Soit:

$$N_t^a(i) = \begin{cases} X_t^a(i) - X_t^{a+1}(i) & \text{si } a \neq a_{end} \\ X_t^{a_{end}}(i) & \text{sinon} \end{cases}$$

Nombre de conseillers présents par période

$N_t(i)$ correspond au nombre de conseillers du profil i présents durant la période t . N_t correspond au nombre total de conseillers présents durant cette période.

$$N_t(i) = \sum_{a \in A(i)} N_t^a(i)$$

$$N_t = \sum_{i \in I} N_t(i)$$

Nombre de conseillers en prise d'appels par période

$N_t^{appels}(i)$ correspond au nombre de conseillers du profil i en prise d'appels durant la période t . N_t^{appels} correspond au nombre total de conseillers en prise d'appels durant cette période.

$$N_t^{appels}(i) = \sum_{a \in A_{appels}(i)} N_t^a(i)$$

$$N_t^{appels} = \sum_{i \in I} N_t^{appels}(i)$$

III.3.5 Les contraintes

Dans ce qui suit nous présenterons la formulation des différentes contraintes à prendre en compte. Elles sont de deux types :

- **Les contraintes de temps de travail.** En plus des contraintes de fenêtre de temps et de durée, nous intégrons également les contraintes permettant de garantir la cohérence des variables de décision telles que l'intégrité des variables ou la succession des activités.
- **Les contraintes d'adéquation entre charge et capacité.** Dans ce qui suit, elles correspondent aux contraintes de couverture de la courbe de charge à traiter.

III.3.5.1 Contraintes de temps de travail

a) Contraintes d'intégrités

Les variables de décision sont des entiers positifs ou nuls car elles représentent un nombre de conseillers.

$$X_t^a(i) \in \mathbb{N} \quad \forall i \in I, \forall a \in A(i), \forall t = \overline{T_0, T_{end}}$$

b) Contraintes de succession des activités

Comme cité plus haut, l'ordre de succession des activités est défini par profil. Afin de garantir cet ordre dans le modèle mathématique il est nécessaire d'introduire, pour chaque profil i , les contraintes suivantes :

$$X_t^a(i) \geq X_t^{a+1}(i) \quad \forall i \in I, \forall a \in A(i) - \{a_{end}\}, \forall t = \overline{T_0, T_{end}}$$

Ces contraintes découlent du fait que si, à un moment donné, un conseiller est dans l'activité $a+1$ alors il est nécessairement passé par l'activité a . Ainsi, à chaque instant t le nombre de conseillers dans l'activité a ou l'ayant déjà effectuée ($X_t^a(i)$) est forcément supérieur au nombre de conseillers dans l'activité $a+1$ ou l'ayant déjà effectuée ($X_t^{a+1}(i)$).

c) *Contraintes imposant des variables de cumul croissantes*

Les variables de décision $X_t^a(i)$ représentent le nombre de conseillers du profil i ayant commencé l'activité a avant la date t . C'est une variable de cumul du nombre de conseiller qui est par définition croissante. Ceci se traduit par les contraintes suivantes :

$$X_{t+1}^a(i) \geq X_t^a(i) \quad \forall i \in I, \forall a \in A(i), \forall t = \overline{T_0, T_{end} - 1}$$

d) *Contrainte de temps de travail maximal et minimal*

$l_{a,a'}(i)$ et $L_{a,a'}(i)$ étant les durées minimale et maximale séparant les débuts des activités a et a' pour le profil i . Ces contraintes de temps de travail seront exprimées comme suit :

$$\begin{cases} X_{t-l_{a,a'}(i)}^a(i) \geq X_t^{a'}(i) & l_{a,a'}(i) \leq t \leq T_{end} \\ X_t^{a'}(i) \geq X_{t-L_{a,a'}(i)}^a(i) & L_{a,a'}(i) \leq t \leq T_{end} \end{cases}$$

Dans le cas où il s'agit de contraindre la durée d'une activité particulière a , il suffit de remplacer dans les inéquations précédentes a' par $a+1$.

Pour illustrer le principe sous-jacent derrière ces contraintes, prenons l'exemple simplifié décrit dans la Figure III-12 d'un centre d'appels avec un profil et une seule activité a (en plus de l'activité fictive de fin de journée de travail a_{end}).

La durée minimale de l'activité a est l (dans l'exemple $l=2$). A chaque instant le nombre de conseillers qui finissent l'activité a (donc qui commencent l'activité a_{end}) est nécessairement inférieur ou égal au nombre de conseillers qui l'ont commencée l périodes avant. De la même manière, si la durée maximale est L (dans l'exemple $L=6$) alors le nombre de conseillers qui finissent l'activité a est supérieur ou égal au nombre de conseillers qui l'ont commencée L périodes avant. De ce fait, la courbe représentant $X_t^{a_{end}}$ ne peut être qu'entre les courbes X_{t-l}^a et X_{t-L}^a

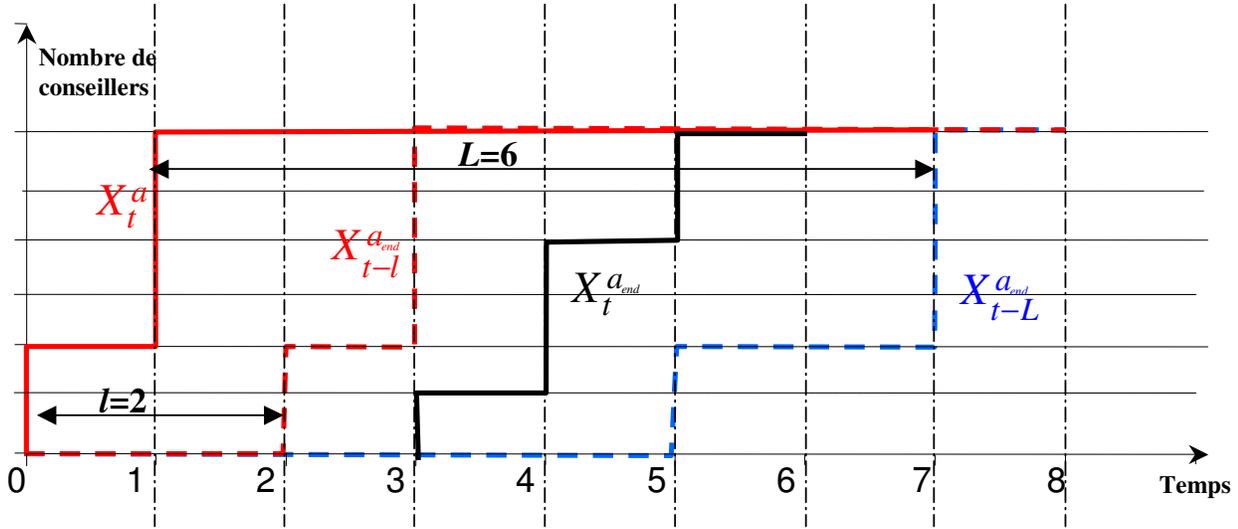


Figure III-12: Illustration des contraintes de durée de travail maximale et minimale

e) Contraintes de fenêtre de temps de travail

$[T_{\min}^a(i), T_{\max}^a(i)]$ correspond à la fenêtre de temps durant laquelle l'activité a doit commencer pour les conseillers du profil i . Cette contrainte s'exprime de la manière suivante :

$$\begin{cases} X_{T_{\min}^a(i)-1}^a(i) = 0 \text{ avec } T_{\min}^a(i) \neq 0 \\ X_{T_{\max}^a(i)}^a(i) = X_{T_{\max}^a(i)}^{a_{end}}(i) = N(i) \end{cases}$$

La première contrainte impose qu'aucun conseiller du profil i n'ait commencé l'activité a avant $T_{\min}^a(i)$. La seconde contrainte impose que tous les conseillers du profil i ont commencé l'activité a avant $T_{\max}^a(i)$.

f) Quelques remarques concernant les contraintes de temps de travail

Activités facultatives et activités obligatoires

Le fait d'imposer un ordre sur les activités d'un profil ne signifie pas nécessairement que chaque conseiller doit effectuer toutes les activités. En effet, si un conseiller "saute" l'activité a ceci se traduit par le fait que $X_t^a(i) = X_t^{a+1}(i)$. Une activité qui n'est pas effectuée est une activité que le conseiller réalise mais pendant une durée nulle. De ce fait, toutes les activités

sont « par défaut » **facultatives**. Si l'on souhaite qu'une activité donnée soit **obligatoire**, nous devons lui associer une durée minimale égale à une période de planification.

Chevauchement des fenêtres de temps

Notons que les contraintes de fenêtres de temps peuvent se chevaucher. Contrairement à (Bechtold et Jacobs 1990), il n'y a pas de restrictions relatives au chevauchement extraordinaire dans notre modèle.

Date ou durée fixe

Si l'on souhaite qu'une activité ait une durée fixe (par exemple, la pause déjeuner qui dure 1 heure exactement), il suffit que les durées maximum et minimum qui lui sont associées soient égales.

De la même façon, pour donner à une activité une date de début fixe, la fenêtre de temps doit être réduite à une seule date

De façon générale, si l'on associe à chaque activité une durée fixe et des dates de début fixes alors on obtient un modèle avec des **vacations explicites**.

Contrainte de fenêtre de temps pour la fin d'une activité

Les contraintes de fenêtre de temps ont été exprimées sur la date de début d'une activité. Notons que pour associer une telle contrainte sur la date de fin d'une activité, il suffit de la définir sur le début de l'activité qui lui succède.

III.3.5.2 Contraintes de couverture

Comme nous l'avons indiqué plus haut, la courbe de charge ou contrainte de couverture correspond au nombre de conseillers minimum par période nécessaires pour couvrir la charge d'appels.

Notons D_t^{appels} le nombre de conseillers nécessaires en prise d'appels durant la période t . Les contraintes de couverture peuvent se traduire comme suit :

$$N_t^{appels} \geq D_t^{appels} \quad \forall t = \overline{T_0, T_{end}}$$

III.3.6 Fonction objectif - Minimisation du coût

L'objectif est de minimiser le coût de dimensionnement. Deux composantes entrent dans la constitution de ce coût :

- le coût de recrutement que l'on notera C_i et qui correspond au coût d'affectation d'un conseiller au profil i .
- le coût horaire que l'on notera c_i et qui correspond au coût de travail par période d'un conseiller du profil i .

Ainsi, la fonction de coût à minimiser s'exprime de la manière suivante :

$$\text{Min } Z = \sum_{i \in I} \left(C_i \cdot N(i) + c_i \cdot \sum_{t=T_0}^{T_{end}} N_t(i) \right)$$

III.3.7 Récapitulatif du modèle

Le problème de construction de vacances traduit dans le paradigme de chaîne d'activités peut donc être formulé par le programme linéaire en nombre entiers suivant :

Objectif

$$MinZ = \sum_{i \in I} \left(C_i \cdot N_i + c_i \cdot \sum_{t=T_0}^{T_{end}} N_t(i) \right) \quad \text{III-1}$$

Contraintes d'intégrité

$$X_t^a(i) \in \mathbb{N} \quad \forall i \in I, \forall a \in A(i), \forall t = \overline{T_0, T_{end}} \quad \text{III-2}$$

Contraintes de temps maximum et minimum de travail

$$\begin{cases} X_{t-l_{a,a'}}^a(i) \geq X_t^{a'}(i) & l_{a,a'}(i) \leq t \leq T_{end} \\ X_t^{a'}(i) \geq X_{t-L_{a,a'}}^a(i) & L_{a,a'}(i) \leq t \leq T_{end} \end{cases} \quad \text{III-3}$$

Contraintes de fenêtres de temps

$$\begin{cases} X_{T_{min}^a(i)-1}^a(i) = 0 \text{ avec } T_{min}^a(i) \neq 0 \\ X_{T_{max}^a(i)}^a(i) = X_{T_{end}}^a(i) = N(i) \end{cases} \quad \text{III-4}$$

Contraintes de successions d'activités

$$X_t^a(i) \geq X_t^{a+1}(i) \quad \forall i \in I, \forall a \in A(i) - \{a_{end}\}, \forall t = \overline{T_0, T_{end}} \quad \text{III-5}$$

Variables de cumul croissantes

$$X_{t+1}^a(i) \geq X_t^a(i) \quad \forall i \in I, \forall a \in A(i), \forall t = \overline{T_0, T_{end} - 1} \quad \text{III-6}$$

Contraintes de couverture

$$N_t^{appels} \geq D_t^{appels} \quad \forall t = \overline{T_0, T_{end}} \quad \text{III-7}$$

avec

$$N_t^{appels} = \sum_{i \in I, a \in A_{appels}(i)} N_t^a(i) \text{ et } N_t^a(i) = \begin{cases} X_t^a(i) - X_t^{a+1}(i) & \text{si } a \neq a_f \\ X_t^{a_{end}}(i) & \text{sinon} \end{cases} \quad \text{III-8}$$

III.4 Analyse des propriétés du modèle proposé

III.4.1 Equivalence entre solution implicite et solution individuelle

Dans un modèle explicite, comme celui de Dantzig, les variables de décisions correspondent au nombre de conseillers affectés aux différentes vacances. A chaque vacation correspond une date de début ainsi qu'une date de fin et éventuellement des dates de début et de fin de pause. Dans ce cas, la solution du modèle peut être déclinée immédiatement en solution individuelle. Ainsi, chaque conseiller de clientèle aura l'emploi du temps qui correspond à celui de la vacation à laquelle il est affecté. La faisabilité d'une solution d'un modèle explicite est donc par définition immédiate.

Dans le cas d'un modèle implicite, par définition, les variables de décision ne sont pas associées aux vacances. De ce fait, le passage de la solution du modèle à la détermination de l'emploi du temps de chaque conseiller n'est pas trivial. Le fait qu'il existe une solution réalisable du problème implicite ne signifie donc pas forcément qu'elle est réalisable à l'échelle individuelle des conseillers. La faisabilité d'une solution d'un modèle implicite ne garantit donc pas a priori sa faisabilité pour le problème de construction de vacation.

L'objet de cette partie, est de présenter une règle simple permettant de décliner la solution du problème implicite en solution individuelle et de démontrer que son utilisation garantit que toute solution réalisable du problème implicite l'est également à l'échelle individuelle. L'application de cette règle nous permet ainsi de montrer que toute solution de notre modèle est une solution du problème de construction de vacation qu'il traduit.

a) Règle premier arrivé - premier parti

Considérons l'exemple où l'on a un seul profil avec une seule pause obligatoire. La Figure III-13 représente une solution du modèle implicite. X_i^1 et X_i^4 correspondent au début et à la fin de la journée de travail. X_i^2 et X_i^3 correspondent au début et à la fin de la pause.

La règle premier arrivé - premier parti permet de déterminer une solution individuelle (c'est à dire explicite) à partir de la solution implicite. Le principe de la règle est le suivant : le premier conseiller à commencer une activité doit être le premier à la terminer.

On démontre qu'en appliquant cette règle, toute solution qui vérifie les contraintes de temps de travail pour le modèle implicite les vérifie également au niveau individuel des conseillers.

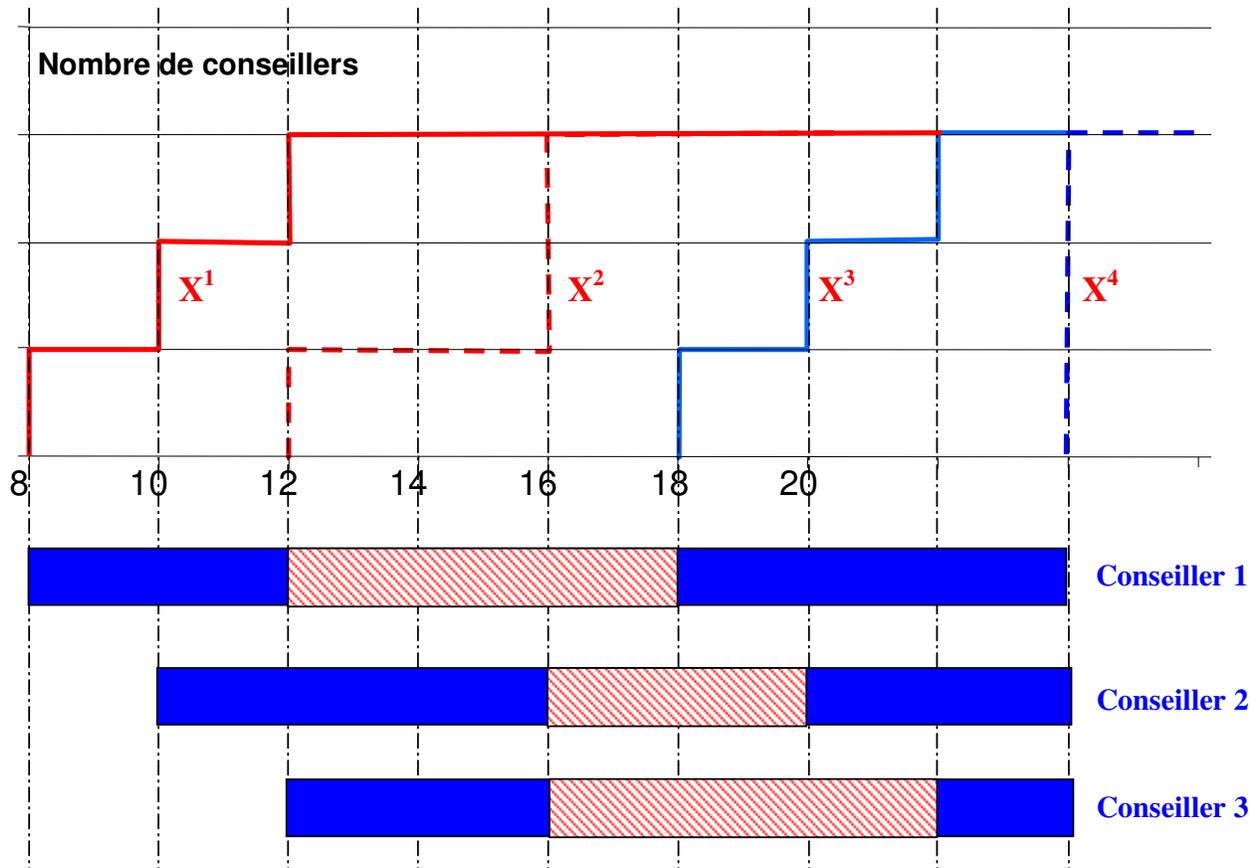


Figure III-13: Illustration de la règle premier arrivé - premier parti

Théorème :

La modélisation implicite du problème de construction de vacation issue du paradigme de chaîne d'activité est équivalente à la modélisation explicite de Dantzig. Toute solution du modèle implicite peut être déclinée en solution individuelle à l'aide de la règle premier arrivé - premier parti.

Démonstration

Introduisons la variable de décision suivante pour traduire les solutions explicites :

$$X_t^a(i, k) = \begin{cases} 1 & \text{si le conseiller } k \text{ du profil } i \text{ a commencé l'activité } a \text{ durant la période } t \text{ ou avant} \\ 0 & \text{sinon} \end{cases}$$

Cette variable de décision correspond à celle que l'on aurait à partir de notre modèle de paradigme de chaîne d'activité où chaque profil n'aurait qu'un seul conseiller.

Les deux variables de décisions sont liées par la relation suivante :

$$X_i^a(i) = \sum_{k=1}^{X_{T_{end}}^{a_{end}}(i)} X_i^a(i, k) \quad \text{III-9}$$

Comme nous pouvons le remarquer de la formulation du modèle (voir récapitulatif III.3.7), les contraintes de temps de travail (contraintes III-3 à III-6) sont de la forme suivante :

$$X_i^a(i) \leq X_i^{a'}(i)$$

Notons que l'égalité dans la contrainte III-4 peut se ramener à une double inégalité de la forme ci-dessus.

Pour prouver le théorème, nous devons démontrer que si une contrainte de temps de travail est respectée dans le modèle implicite, elle l'est aussi dans le modèle explicite et vice versa soit:

$$X_i^a(i) \leq X_i^{a'}(i) \Leftrightarrow X_i^a(i, k) \leq X_i^{a'}(i, k) \quad \forall k = 1, \overline{X_{T_{end}}^{a_{end}}(i)}$$

$$\mathbf{a) \text{ Explicite} \Rightarrow \text{implicite} : } X_i^a(i, k) \leq X_i^{a'}(i, k) \quad \forall k = 1, \overline{X_{T_{end}}^{a_{end}}(i)} \Rightarrow X_i^a(i) \leq X_i^{a'}(i)$$

Si $X_i^a(i, k) \leq X_i^{a'}(i, k) \quad \forall k = 1, \overline{X_{T_{end}}^{a_{end}}(i)}$ alors en sommant on obtient

$$\sum_{k=1}^{X_{T_{end}}^{a_{end}}(i)} X_i^a(i, k) \leq \sum_{k=1}^{X_{T_{end}}^{a_{end}}(i)} X_i^{a'}(i, k)$$

De l'égalité III-9 on peut déduire que : $X_i^a(i) \leq X_i^{a'}(i)$

La première implication est donc démontrée

$$\mathbf{b) \text{ Implicite} \Rightarrow \text{explicite} : } X_i^a(i) \leq X_i^{a'}(i) \Rightarrow X_i^a(i, k) \leq X_i^{a'}(i, k) \quad \forall k = 1, \overline{X_{T_{end}}^{a_{end}}(i)}$$

Sans perdre en généralité, nous supposons que les conseillers sont ordonnés en fonction de la période de début de travail. Ainsi, pour un profil donné, le conseiller $k=1$ est le premier

à travailler, $k=2$ est le second, etc. Si a_1 est la première activité effectuée durant la journée pour un profil i on aura :

$$X_i^{a_1}(i, k) = \begin{cases} 1 & \text{si } X_i^{a_1}(i) \geq k \\ 0 & \text{sinon} \end{cases} \quad \forall i \in I, \forall t = \overline{T_0, T_{end}}$$

Etant donné qu'au sein d'un même profil tous les conseillers effectuent les mêmes activités dans le même ordre, l'application de la règle premier arrivé - premier parti conduira à avoir le même ordre de passage des conseillers sur toutes les activités. La relation ci-dessus peut donc être généralisée à toutes les activités. Soit :

$$X_i^a(i, k) = \begin{cases} 1 & \text{si } X_i^a(i) \geq k \\ 0 & \text{sinon} \end{cases} \quad \forall i \in I, a \in A(i), \forall t = \overline{T_0, T_{end}} \quad \text{III-10}$$

Supposons que le premier terme de l'implication est vérifié. Soit

$$X_i^a(i) \leq X_i^{a'}(i) \text{ nous devons vérifier que } X_i^a(i, k) \leq X_i^{a'}(i, k)$$

Si l'on considère un conseiller k du profil i , alors on aura deux cas de figure :

1^{er} cas : si $k \geq X_i^a(i)$ alors d'après III-10

$$X_i^a(i, k) = 0 \Rightarrow X_i^a(i, k) \leq X_i^{a'}(i, k) \text{ car } X_i^{a'}(i, k) \text{ est positif ou nul}$$

2nd cas : si $k < X_i^a(i) \leq X_i^{a'}(i)$ alors d'après III-10

$$X_i^{a'}(i, k) = X_i^a(i, k) = 1 \Rightarrow X_i^a(i, k) \leq X_i^{a'}(i, k) \text{ c.q.f.d}$$

III.4.2 Analyse de la difficulté du problème

Dans le cas général, le problème de *shift-scheduling* est NP-difficile au sens fort (Garey et Johnson 1979; Bartholdi 1981). Dans ce qui suit, nous montrons que notre modélisation présente une structure particulière : la totale unimodularité. Cette propriété nous permettra de mettre en évidence deux éléments :

- Analyser le lien entre contraintes de couverture et la difficulté du problème de *shift-scheduling*.
- Identifier des cas particuliers pour lesquels le problème de *shift-scheduling* peut être résolu en un temps polynomial.

III.4.2.1 La totale unimodularité de la matrice des contraintes de temps de travail

Nous reprenons la définition d'une matrice totalement unimodulaire indiquée dans (Ghouila-Houri 1962).

Définition :

On dit qu'une matrice est totalement unimodulaire si le déterminant de toutes les sous-matrices carrées est à égal à 0, +1 ou -1 ; il faut en particulier que tous les termes de cette matrice soient égaux à 0, +1 ou -1.

Cette propriété est intéressante car si un problème peut être formulé en un programme linéaire en nombres entiers dont la matrice de contrainte est totalement unimodulaire alors la solution optimale obtenue est nécessairement entière même si les contraintes d'intégrité sont relaxées. Le temps de résolution d'un PLNE dont la matrice de contrainte est totalement unimodulaire est donc polynomial.

Théorème :

La matrice de contrainte représentant les contraintes de temps de travail issues du paradigme de chaîne d'activités est totalement unimodulaire

a) Démonstration du théorème

Avant de démontrer le théorème, intéressons-nous à la forme particulière des contraintes de temps de travail.

Ces contraintes (III-3 à III-6) sont de la forme suivante :

$$\begin{aligned}x &\leq y \\x &\leq a \\x &\geq b\end{aligned}$$

Remarque : Les deux dernières formes de contraintes sont obtenues en remplaçant les variables des égalités III-4 dans les autres contraintes d'inégalité.

Cette propriété peut être illustrée graphiquement (voir Figure III-14). Considérons le cas où l'on a deux variables de décision x et y et les contraintes suivantes :

$$\begin{aligned}y &\leq x \\y &\leq 2 \\x &\leq 4\end{aligned}$$

Graphiquement, les droites correspondant à ces contraintes forment un angle de 90° , 0° ou 45° avec l'axe des abscisses. Dans ce cas, tous les sommets constitués par les intersections des contraintes sont sur des points entiers. Etant donné que, pour un programme linéaire, les solutions sont sur les sommets, on comprend aisément qu'il n'est pas nécessaire, dans ce cas, d'imposer des contraintes d'intégrité.

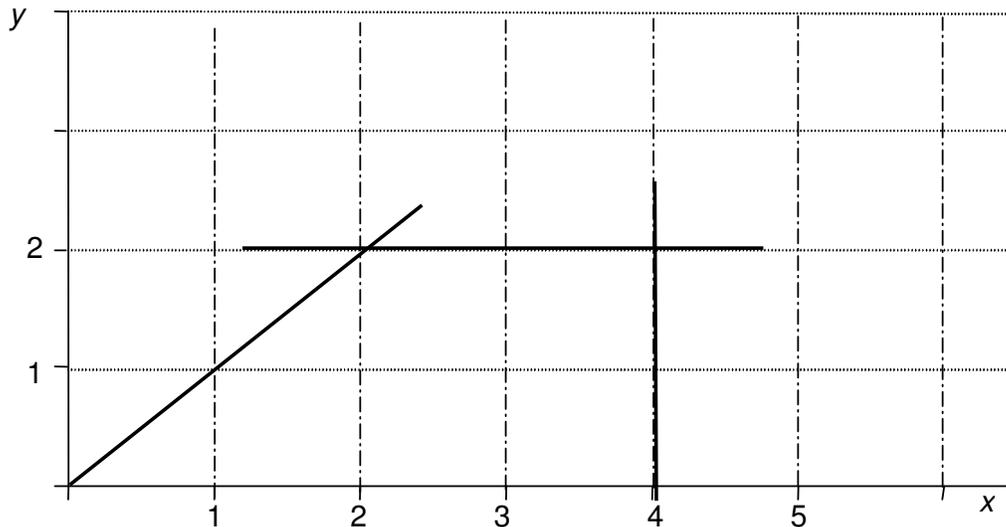


Figure III-14: Intersection des contraintes sur des points entiers

De manière plus générale et plus rigoureuse qu'une représentation graphique, abordons à présent la démonstration du théorème. Pour ce faire nous allons recourir au théorème fondamental de Ghouila-Houri (Ghouila-Houri 1964)

Théorème fondamental de Ghouila-Houri

Soit une matrice $M = m_j^i$ ($i \in I; j \in J$) dont les termes sont égaux à 0, +1 ou -1. Les deux propositions suivantes sont équivalentes :

- M est totalement unimodulaire
- Tout ensemble non vide $A \subset I$ peut être partitionné en deux ensembles A_1 et A_2 tel que :

$$\sum_{i \in A_1} m_j^i - \sum_{i \in A_2} m_j^i = 0, +1 \text{ ou } -1 \quad \forall j \in J$$

Le principe de notre démonstration repose sur ce théorème fondamental. Multiplions chaque ligne de notre matrice de contrainte par un coefficient +1 ou -1. Nous démontrons

qu'il existe, compte tenu de la structure particulière de notre matrice de contraintes, une affectation de ces coefficients de telle sorte que la somme sur les lignes donne 0, +1 ou -1.

Cas des contraintes de la forme $x \leq y$

Pour l'instant, supposons que les contraintes sont toutes de la forme $x \leq y$. Nous étendrons la démonstration aux contraintes $x \leq a$ et $x \geq b$ dans un second temps.

Ces contraintes se traduisent dans la forme matricielle suivante

$$\begin{bmatrix} 0 & \dots & +1 & \dots & -1 & 0 \\ \dots & -1 & +1 & 0 & \dots & 0 \\ \dots & +1 & 0 & \dots & -1 & \dots \end{bmatrix}$$

C'est une matrice où chaque ligne contient exactement deux éléments non nuls +1 et -1.

Le fait que les lignes contiennent exactement deux valeurs non nulles nous permet de traduire cette matrice en un graphe équivalent que nous définissons de la manière suivante :

- Les sommets du graphe représentent les colonnes
- Les arcs représentent les lignes
- Deux sommets sont reliés par un arc si les colonnes qui leur correspondent contiennent +1 et -1 sur une même ligne
- L'arc est orienté du sommet ayant la valeur +1 au sommet ayant la valeur -1

Le principe de cette traduction en graphe équivalent est illustré dans la Figure III-15

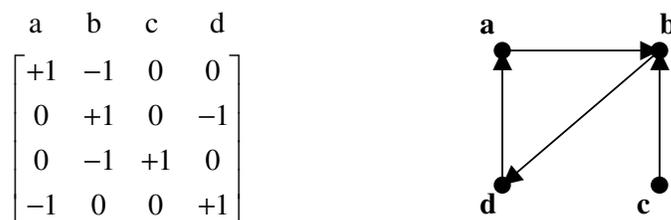


Figure III-15: Traduction de la matrice de contrainte en graphe équivalent

Nous pouvons faire deux remarques :

- i) Par construction, les arcs sortants d'un sommet correspondent à un +1 dans la matrice de contrainte et les sommets entrants correspondent à un -1. Ainsi, la somme des valeurs sur une colonne est égale à la différence pour le sommet correspondant entre le nombre d'arcs sortants et le nombre d'arcs entrants.
- ii) Multiplier une ligne par -1 conduit à inverser le sens de l'arc correspondant dans le graphe équivalent.

A partir de ces deux remarques, nous pouvons déduire que les deux propositions suivantes sont équivalentes :

- Il existe une orientation des arcs du graphe équivalent de telle sorte que la différence entre le nombre d'arcs sortants et le nombre d'arcs entrants est égale à 0, +1 ou -1
- Il existe une affectation de coefficients multiplicatifs +1 ou -1 de telle sorte que la somme sur les lignes de la matrice de contrainte donne 0, +1 ou -1.

Nous aboutissons donc à la propriété suivante :

Propriété 1 :

Une matrice de contraintes de la forme $x \leq y$ est totalement unimodulaire s'il existe une orientation des arcs du graphe équivalent de telle sorte que la différence entre le nombre d'arcs sortants et entrants pour chaque somme est égale à 0, +1 ou -1

Il suffit donc de construire une orientation du graphe équivalent qui vérifie cette propriété pour démontrer notre théorème. Pour ce faire, considérons les deux cas possibles pour un graphe :

- graphe sans cycle
- graphe avec cycle

a) cas d'un graphe sans cycle

Un graphe sans cycle est constitué d'un ou plusieurs arbres. Sans perdre en généralité, nous supposons que notre graphe équivalent est constitué d'une seule composante connexe et donc d'un seul arbre. En effet, pour que la propriété 1 soit vérifiée sur un graphe constitué de plusieurs composantes connexes, il est nécessaire et suffisant qu'elle soit vérifiée sur chaque composante connexe séparément.

Nous démontrons par récursivité la propriété 1.

Cas d'un arbre avec un seul arc

De manière évidente, la propriété 1 est vérifiée : le graphe est constitué de deux sommets dont l'un possède un arc sortant et l'autre un arc entrant. La différence entre le nombre de sommets sortants et entrants est donc égale à +1 et -1.

Propriété vérifiée pour un arbre de n arcs

Considérons l'arbre $A(n)$ de n arcs qui vérifié la propriété 1. Il existe donc une orientation des arcs de telle sorte que pour chaque sommet la différence entre le nombre de sommets sortants et entrants est égale à 0, +1 ou -1.

Ajoutons un arc supplémentaire a de manière à obtenir un nouvel arbre $A(n+1)$ de $n+1$ arcs. Cet arc ne peut relier deux sommets de $A(n)$ car on aurait alors un cycle. Ainsi deux cas sont possibles :

- a ne relie aucun sommet de $A(n)$. Il n'y a pas de modification sur le nombre d'arcs qui relient chaque sommet. La propriété 1 est donc toujours vérifiée pour l'arbre $A(n+1)$
- a relie un seul sommet de $A(n)$ que nous désignons s . Avant l'ajout de a , la différence entre le nombre de sommets sortants et entrants pour s pouvait avoir 3 valeurs :
 - o 0 : la propriété 1 est toujours vérifiée après l'ajout de a car cette différence est à présent égale à +1 ou -1
 - o +1 : il suffit d'orienter a de telle sorte que s en soit l'origine pour la propriété 1 soit toujours vérifiée (la différence est à présent égale à 0)
 - o -1 : il suffit d'orienter a de telle sorte que s en soit la destination pour la propriété 1 soit toujours vérifiée (la différence est à présent égale à 0)

Ainsi, dans tous les cas, la propriété 1 est toujours vérifiée après l'ajout d'un arc. Notre théorème est donc démontré pour les matrices de contraintes de la forme $x \leq y$ dont le graphe équivalent est sans cycles.

b) cas d'un graphe avec cycle

Par définition, il est possible d'orienter les arcs d'un cycle de manière à obtenir un circuit. Chaque sommet d'un circuit possède un arc entrant et un arc sortant. La différence entre les deux est donc nulle. De ce fait, il est possible de supprimer du graphe équivalent les arcs qui

constituent un cycle (en les orientant pour aboutir à un circuit) et ce sans modifier la différence entre nombre d'arcs sortants et entrants pour tous les sommets.

Ainsi, si on procède par itération en identifiant à chaque étape un cycle et en supprimant les arcs qui le constituent, on finit par aboutir à un graphe sans cycle qui ne modifie pas la différence entre nombre d'arcs sortants et entrants pour chaque sommet. Comme nous avons démontré plus haut qu'un graphe sans cycle vérifiait la propriété 1, nous pouvons donc conclure qu'un graphe avec cycle la vérifie également.

Ceci achève la démonstration de notre théorème pour la matrice de contraintes de la forme $x \leq y$: elle est bien totalement unimodulaire. Introduisons à présent les autres contraintes de la forme $x \leq a$ et $x > b$.

Cas des contraintes sous la forme $x \leq a$ et $x \geq b$

Nous avons démontré que la matrice des contraintes de la forme $x \leq y$ était totalement unimodulaire. Si A désigne l'ensemble des lignes de cette matrice alors d'après le théorème fondamental de Ghouila-Houri cité plus haut, il existe deux ensembles A_1 et $A_2 \subset A$ tels que :

$$S_j = \sum_{i \in A_1} m_j^i - \sum_{i \in A_2} m_j^i = 0, +1 \text{ ou } -1 \quad \forall j \in J$$

Les contraintes de la forme $x \leq a$ (respectivement $x \geq b$) correspondent, en écriture matricielle, à une ligne dont tous les éléments sont nuls sauf un seul qui est égal à 1 (respectivement -1). Notons B (respectivement C) l'ensemble des lignes de la matrice de contraintes de la forme $x \leq a$ (respectivement $x \geq b$). Notons $b_k \in B$ (respectivement $c_k \in C$) la ligne qui correspond à la contrainte dont le k ème élément est égal à 1 (respectivement -1).

Ajoutons itérativement chaque ligne b_k (respectivement c_k) à la somme S_j en la multipliant au préalable par -1 (respectivement +1) si $S_j = +1$ ou 0 et par +1 (respectivement -1) si $S_j = -1$. A chaque étape la somme est toujours égale à 0, +1 ou -1 jusqu'à ce que l'on ait ajouté toutes les lignes. Ainsi le théorème fondamental de Ghouila-Houri est toujours vérifié.

Ceci achève la démonstration de notre théorème.

III.4.2.2 Introduction des contraintes de couverture

Si, à présent, on considère les contraintes de couverture, on voit que, dans le cas général, la contrainte III-7 fait que la matrice des contraintes n'est pas totalement unimodulaire. En effet, elle s'exprime de la manière suivante :

$$\sum_{i \in I, a \in A_{appels}} X_t^a(i) - X_t^{a+1}(i) \geq D_t^{appels} \quad \forall t = \overline{T_0, T_{end}}$$

Dès lors que le nombre de profils et/ou d'activités est supérieur ou égal à 2, il est possible de construire des contre-exemples pour lesquels la matrice n'est plus totalement unimodulaire.

Ce sont donc ces contraintes qui rendent le problème NP-difficile. Dans certains cas particuliers, les contraintes de couvertures ne modifient pas la structure totalement unimodulaire de la matrice de contrainte. Ces cas correspondent aux problèmes de *shift-scheduling* qui peuvent être résolu en temps polynomial.

III.4.2.3 Cas particulier polynomial

Nous proposons dans ce qui suit d'effectuer une analyse sur notre modèle implicite afin d'identifier les conditions permettant d'avoir un problème polynomial.

Comme nous l'avons indiqué plus haut, il suffit de déterminer les conditions pour lesquelles les contraintes de couverture forment une matrice totalement unimodulaire pour que le problème devienne polynomial.

Profil unique, activité unique, pas de pause

Si l'on considère le cas présent où l'on a qu'une seule activité (en plus de l'activité fictive qui correspond à la fin de la journée de travail), un seul profil et pas de pause. La contrainte de couverture devient :

$$X_t^a - X_t^{end} \geq D_t^{appels} \quad \forall t = \overline{T_0, T_{end}}$$

Cette nouvelle formulation fait que, dans ce cas, notre matrice de contraintes devient totalement unimodulaire car de la forme $x \leq y$. Le problème peut donc être résolu en un temps polynomial par un programme linéaire en relaxant les contraintes d'intégrité.

Théorème :

Le problème de *shift scheduling* avec minimisation du coût de recrutement peut être résolu en un temps polynomial dans le cas où :

- il n'y a pas de pause
- il n'y a qu'un seul type d'activité (prise d'appel) et un seul profil
- avec éventuellement une contrainte de durée de travail maximum et minimum
- éventuellement des fenêtres de temps pour le début et la fin d'activité

III.5 Généralisation du modèle issu du paradigme de chaîne d'activité à la multi-compétence

Jusque là nous avons considéré dans notre modèle un seul type de flux : des appels qui peuvent être traités par tous les conseillers. Dans ce qui suit, nous montrons que le modèle de programmation linéaire traduisant le paradigme de chaîne d'activité peut être généralisé au cas où plusieurs flux doivent être traités. Dans ce qui suit nous considérons que les conseillers ont des compétences différentes qui définissent les flux qu'ils peuvent traiter.

Nous distinguerons les cas de figure suivant :

- Plusieurs flux d'appels
- Un mix entre flux d'appels (non stockable) et flux stockable (mail, courrier, etc.)
- Le call blending où les conseillers peuvent basculer d'un dossier stockable à un appel

Il existe une littérature riche abordant la problématique de la multi-compétence dans un centre d'appels. Pour une vue générale voir (Gans, Koole et al. 2003; Wallace et Whitt 2005). Concernant le *shift-scheduling* plus spécifiquement, (Bhulai, Koole et al. 2007) propose une méthode en deux étapes pour résoudre le *shift-scheduling*. Dans la première étape, les contraintes de couvertures sont déterminées. Dans la seconde étape, le problème de *shift-scheduling* est résolu par un modèle de programmation linéaire. Nous pouvons également faire référence à (Cezik et L'Ecuyer 2008) qui intègre la multi-compétence en adaptant l'approche de (Atlason, Epelman et al. 2004) combinant programmation linéaire et simulation.

Les références que nous citons proposent des approches reposant sur le modèle **explicite** de (Dantzig 1954). Dans ce qui suit nous proposons une adaptation de notre modèle afin de garder les avantages d'une représentation **implicite** des vacations.

III.5.1 Plusieurs flux d'appels (synchrone)

Dans ce cas nous avons plusieurs flux d'appels différents qui appartiennent à l'ensemble F_{appels} .

Les chaînes d'activité des profils correspondent à une succession d'activités d'appels et d'activités de pause. A chaque profil est associé un ensemble de compétence qui correspond aux flux qui peuvent être traités par les conseillers. Durant une activité d'appels (qui appartient à $A_{appels}(i)$), tous les conseillers peuvent traiter les flux dont ils ont la compétence.

Nous supposons par ailleurs qu'il n'y pas de système de priorité dans la distribution des appels en fonction de la compétence.

A chaque flux correspond une contrainte de couverture D_t^f qui représente le nombre minimum de conseillers nécessaire pour traiter la charge du flux f durant la période t (s'ils devaient traiter exclusivement ce type d'appel).

Nous introduisons la variable de décision supplémentaire $p_t^f(i)$ qui correspond à la proportion de la charge D_t^f traitée par les conseillers du profil i .

Rappelons que $N_t^{appels}(i)$ correspond au nombre de conseillers du profil i qui effectuent une activité de traitement d'appels durant la période t . Avec :

$$N_t^{appels}(i) = \sum_{a \in A_{appels}} N_t^a(i) \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

Ainsi les contraintes de couverture se traduisent de la manière suivante

$$N_t^{appels}(i) \geq \sum_{f \in F_{appels}} p_t^f(i) \cdot D_t^f \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

Avec

$$\sum_{i \in I} p_t^f(i) = 1 \quad \forall t = \overline{T_0, T_{end}}, \forall f \in F_{appels}$$

Les contraintes liées aux compétences des conseillers peuvent s'exprimer simplement en ajoutant les contraintes suivantes

$$p_t^f(i) = 0 \quad \forall t = \overline{T_0, T_{end}} \text{ si le profil } i \text{ ne peut pas traiter le flux } f.$$

Notons que bien qu'il ne soit pas nécessaire d'ajouter des activités supplémentaires, cette formulation introduit une dimension de complexité supplémentaire liée aux compétences. En effet, même si tous les conseillers ont tous les mêmes contraintes de temps de travail, il faut définir autant de profil qu'il y a de configuration de compétences possibles.

III.5.2 Plusieurs flux stockable (asynchrone) et plusieurs flux d'appels

Les flux stockables sont les flux dont le traitement peut être reporté aux périodes suivantes s'il n'est pas possible de les traiter durant leur période d'arrivée. Il s'agit par

exemple des appels sortant qu'émet le conseiller, les courriers, les emails, etc. On parle également de flux asynchrone par opposition des flux synchrones que sont les appels.

Nous supposons dans cette partie (voir Figure III-16) que certaines activités sont dédiées aux flux stockables et d'autres aux flux d'appels (en plus de celles correspondant à des pauses). Ainsi, un conseiller pourra par exemple alterner entre le traitement d'un courrier puis d'un email mais il ne pourra pas prendre d'appels tant qu'il est dans une activité de traitement de flux stockable.

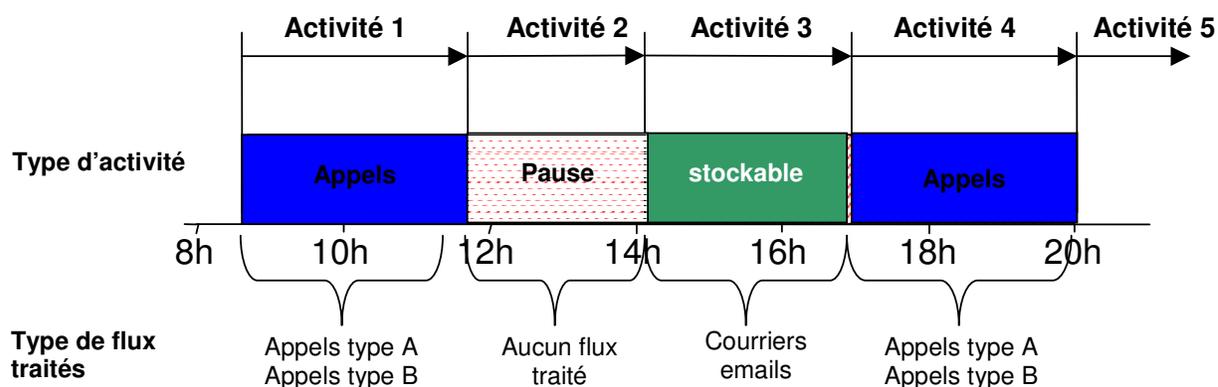


Figure III-16: Illustration du cas avec plusieurs flux stockable et flux d'appels

III.5.2.1 Un seul flux stockable

Pour des raisons de clarté commençons par considérer un seul type de flux stockable, par exemple le courrier.

Soit $D_{courrier_t}$ le nombre de conseillers nécessaires pour traiter le volume de courriers qui arrivent durant la période t .

Nous supposons que l'objectif est de traiter tous les courriers **le jour même**. Ainsi, pour respecter cette contrainte, il suffit que la capacité de service liée aux activités de courrier soit supérieure à la capacité nécessaire au traitement de tous les courriers arrivés sur une journée, soit :

$$\sum_{t=T_0}^{T_{end}} \sum_{i \in I} N_t^{courrier}(i) \geq \sum_{t=T_0}^{T_{end}} D_{courrier_t}$$

III-11

Avec $N_t^{courrier}(i)$ le nombre de conseillers du profil i qui effectuent une activité de traitement de courrier durant la période t . On a ainsi :

$$N_t^{courrier}(i) = \sum_{a \in A_{courrier}} N_t^a(i) \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

La contrainte III-11 permet de garantir que tous les courriers arrivés dans la journée sont traités. En revanche, il est également nécessaire de vérifier que le traitement ne se fait pas avant l'arrivée du courrier. Ainsi, si tous les courriers arrivent l'après midi, le traitement ne peut se faire le matin. Pour cela il est nécessaire de vérifier que tous les courriers qui arrivent après une période T ne seront traités que par des ressources disponibles après cette période T . Ceci se traduit dans le modèle par la généralisation de la contrainte III-11 qui donne l'ensemble des contraintes suivantes :

$$\sum_{t=T}^{T_{end}} \sum_{i \in I} N_t^{courrier}(i) \geq \sum_{t=T}^{T_{end}} D_{courrier}_t \quad \forall T = \overline{T_0, T_{end}} \quad \text{III-12}$$

III.5.2.2 Plusieurs flux stockable

Considérons, à présent que plusieurs flux stockables peuvent être traités. A l'instar de ce qui a été fait dans la partie III.5.1, nous définissons $F_{stockable}$ l'ensemble des flux stockables. A chaque profil est associé un ensemble de compétence qui correspond aux flux qui peuvent être traités par les conseillers. Tous les conseillers d'un même profil peuvent traiter les mêmes flux.

Soit $A_{stockable}$ l'ensemble des activités correspondant au traitement des flux stockables. Avec $N_t^{stockable}(i)$ le nombre de conseillers du profil i qui effectuent une activité de traitement de flux stockable durant la période t . On a ainsi :

$$N_t^{stockable}(i) = \sum_{a \in A_{stockable}} N_t^a(i) \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

La charge est représentée par $D_{stockable}_t^f$, le nombre de conseillers nécessaires pour traiter le volume de flux stockable f qui arrive durant la période t .

La contrainte d'adéquation entre charge et capacité III-12 peut se généraliser comme suit au cas de plusieurs flux stockable :

$$\sum_{t=T}^{T_{end}} N_t^{stockable}(i) \geq \sum_{f \in F_{stockable}} \sum_{t=T}^{T_{end}} p_t^f(i) \cdot D_{stockable}_t^f \quad \forall T = \overline{T_0, T_{end}}, \forall i \in I \quad \text{III-13}$$

Avec

$$\sum_{i \in I} p_t^f(i) = 1 \quad \forall t = \overline{T_0, T_{end}}, \forall f \in F_{stockable}$$

Comme précédemment, les compétences des conseillers peuvent être prises en compte par les contraintes suivantes :

$$p_t^f(i) = 0 \quad \forall t = \overline{T_0, T_{end}} \text{ si le profil } i \text{ ne peut pas traiter le flux } f.$$

III.5.3 Le cas du *Call blending*

Dans ce qui précède, nous avons considéré qu'un conseiller ne pouvait basculer du traitement d'un flux stockable à un appel au sein d'une même activité (et a fortiori au sein d'une même période). Le *call blending* permet de lever cette restriction (Reynolds 2003).

Le *call blending* est une fonctionnalité offerte par le mécanisme de routage du centre d'appels qui permet au conseiller de traiter des flux stockables et des appels lors d'une même période. Ainsi, un conseiller pourra alterner d'un email à un appel en temps réel selon les règles de routage mises en place. Ce cas général avec des activités dédiées aux appels et aux flux stockables est illustré dans la Figure III-17.

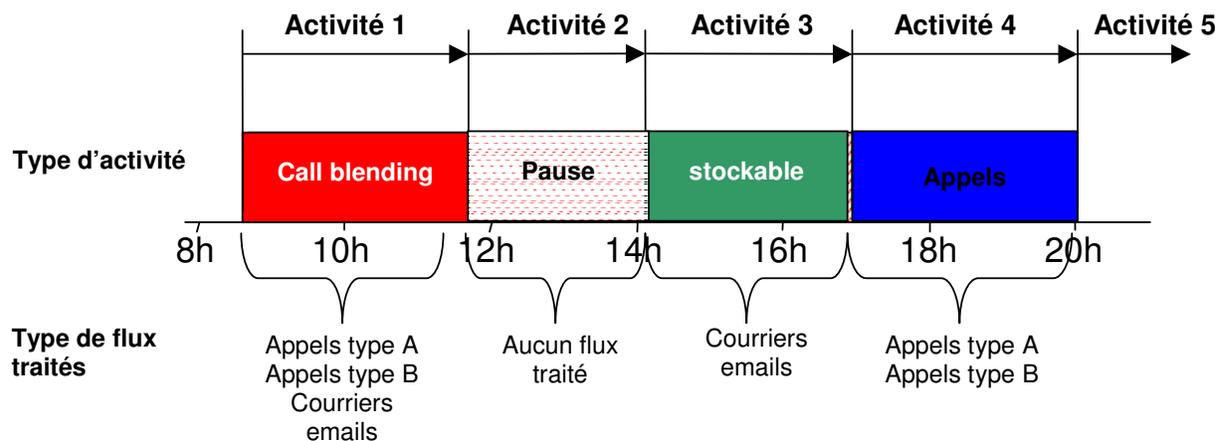


Figure III-17: Illustration du cas avec plusieurs flux stockable, flux d'appels et call-blending

Etant donné que la qualité de service attendue pour les appels est bien supérieure à celle des flux stockable (de l'ordre de la minute pour les appels et quelques jours pour les courriers), les mécanismes de routages sont paramétrés de sorte que le traitement des appels soit prioritaire sur tous les flux stockable.

Nous supposons que la priorité des appels est préemptive. C'est-à-dire que si un appel arrive alors qu'un conseiller est en cours de traitement d'un courrier, ce dernier interrompra son traitement pour prendre l'appel. Nous supposons, de plus, que cette interruption ne génère pas au final un temps de traitement plus long du courrier en question (le temps nécessaire à la reprise d'un dossier stockable interrompu par un appel est supposé nul).

Nous définissons un nouvel ensemble d'activité supplémentaire A_{mix} qui correspond aux activités au cours desquels un conseiller peut prendre indifféremment des appels ou des flux stockables. Avec $N_t^{mix}(i)$ le nombre de conseillers du profil i qui effectuent une activité de ce type durant la période t . On a ainsi :

$$N_t^{mix}(i) = \sum_{a \in A_{mix}} N_t^a(i) \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

Nous introduisons la variable de décision supplémentaire $pmix_t^f(i)$ qui correspond à la proportion de la charge D_t^f traitée par les conseillers du profil i en train d'effectuer une activité de *call-blending* (ie appartenant à A_{mix}). Nous gardons la variable de décisions $p_t^f(i)$ qui correspond proportion de la charge D_t^f traitée par les conseillers du profil i en train d'effectuer une activité d'appels si $f \in F_{appels}$ ou une activité stockable si $f \in F_{stockable}$

La première contrainte d'adéquation entre charge et capacité est de vérifier qu'il y a suffisamment de capacité pour prendre les appels. Soit :

$$N_t^{appels}(i) \geq \sum_{f \in F_{appels}} p_t^f(i) \cdot D_t^f \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

$$N_t^{mix}(i) \geq \sum_{f \in F_{appels}} pmix_t^f(i) \cdot D_t^f \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I$$

Avec

$$\sum_{i \in I} pmix_t^f(i) + p_t^f(i) = 1 \quad \forall t = \overline{T_0, T_{end}}, \forall f \in F_{appels}$$

Etant donné que les appels sont traités de manière prioritaire, la capacité disponible pour traiter des flux stockable lors des activités en *call-blending* pour le profil i durant la période t s'exprime comme suit :

$$N_t^{mix}(i) - \sum_{f \in F_{appels}} pmix_t^f(i) \cdot D_t^f$$

La contrainte d'adéquation entre charge et capacité pour les flux stockables peut donc se traduire comme suit :

$$\sum_{t=T}^{T_{end}} N_t^{stockable}(i) \geq \sum_{f \in F_{stockable}} \sum_{t=T}^{T_{end}} p_t^f(i) \cdot D_{stockable}^f \quad \forall T = \overline{T_0, T_{end}}, \forall i \in I$$

$$\sum_{t=T}^{T_{end}} \left(N_t^{mix}(i) - \sum_{f \in F_{appels}} pmix_t^f(i) \cdot D_t^f \right) \geq \sum_{f \in F_{stockable}} \sum_{t=T}^{T_{end}} pmix_t^f(i) \cdot D_{stockable}^f \quad \forall T = \overline{T_0, T_{end}}, \forall i \in I$$

Avec

$$\sum_{i \in I} pmix_t^f(i) + p_t^f(i) = 1 \quad \forall t = \overline{T_0, T_{end}}, \forall f \in F_{appels}$$

III.5.4 Equivalence implicite-explicite du modèle généralisé

La généralisation de notre modèle n'a pas modifié les contraintes de temps de travail. Ainsi, la propriété démontrée dans III.4.1 reste valable dans le cas général. Ceci garantit que la propriété d'équivalence entre modèle implicite et explicite est également vérifiée par notre modèle général. Il est donc toujours possible à partir de ce modèle de trouver une solution individuelle.

III.6 Positionnement du modèle par rapport à la littérature

Afin de positionner la contribution de notre paradigme par rapport aux travaux antérieurs, reprenons dans le tableau Tableau III-2 les principaux éléments de synthèse cités à la fin de l'état de l'art (voir Tableau I-1).

Modèle	Vacation hors pause implicite	Une seule pause implicite	Plusieurs pauses implicites	Multi-flux et multi-compétence	Flux stockable	Activités facultatives et obligatoires
Dantzig, 1954	-	-	-	-	-	-
Segal, 1974	-	-	-	-	-	-
Moondra, 1976	✓	-	-	-	-	-
Bechtold, 1990	-	✓	-	-	-	-
Thompson, 1995	✓	✓	-	-	-	-
Aykin, 1996	-	✓	✓	-	-	-
PLNE issu du paradigme	✓	✓	✓	✓	✓	✓

Tableau III-2 : Apport du modèle de PLNE issu du paradigme de chaîne d'activités

Comme nous l'avons indiqué, notre modèle est dans la continuité des précédents modèles qui représentent de manière implicite les pauses et les vacations.

Il va au-delà des précédents travaux dans la mesure où il est totalement implicite en ce qui concerne les vacations (dates de début et de fin et durées) ainsi que les pauses (nombre de pauses et durées quelconque). Il permet également de prendre en compte des éléments supplémentaires tels que les flux stockables, la multi-compétence/multi-flux ainsi que la possibilité de distinguer entre activités facultatives et obligatoires.

III.7 Résolution du problème de dimensionnement - Qualité de la relaxation linéaire

Comme nous l'avons vu dans le second chapitre, le dimensionnement se fait à long terme. Aussi, les données à ce stade sont incertaines. La capacité de service résultant des décisions de recrutement du dimensionnement devra être ajustée lors des étapes ultérieures (par recours à la sous-traitance, aux intérimaires, aux CDD). Ainsi, du fait de cette incertitude, la solution optimale du problème de dimensionnement a un intérêt limité. Ce qui est important c'est d'obtenir une solution robuste qui servira de base aux étapes ultérieures.

Pour cette raison, nous ne chercherons pas dans ce qui suit à résoudre de manière optimale le problème de dimensionnement. Le problème étant NP-difficile, le temps de résolution serait exponentiel. Nous proposons donc de résoudre le programme linéaire en relaxant les contraintes d'intégrité puis d'arrondir la solution ainsi obtenue. Ceci permet d'utiliser des méthodes de résolution polynomiales. La rapidité d'exécution est un point important lorsqu'il s'agit de proposer aux utilisateurs des outils interactifs autorisant des essais successifs.

III.7.1 Faisabilité de la solution arrondie

Soient $X^* = (X_i^a(i))$ la solution optimale du problème et $\tilde{X}^* = (\tilde{X}_i^a(i))$ la solution optimale du problème en relaxant les contraintes d'intégrité.

Nous démontrons que la partie entière de la solution relaxée noté $\lfloor \tilde{X}^* \rfloor = (\lfloor \tilde{X}_i^a(i) \rfloor)$ respecte toutes les contraintes de temps de travail.

Proposition :

si une solution $\tilde{X}^* = (\tilde{X}_i^a(i))$ respecte les contraintes de temps de travail associées au paradigme de chaîne d'activité alors la partie entière de cette solution les respecte aussi.

Démonstration

Toutes les contraintes de temps de travail (cf. III.3.5.1) sont de la forme

$$x \leq y \text{ ou } x = y \text{ avec } x \geq 0, y \geq 0$$

Le cas $x=y$ est trivial car $x = y \Rightarrow \lfloor x \rfloor = \lfloor y \rfloor$. Ainsi pour démontrer que la proposition ci-dessus est correcte, il suffit de démontrer la proposition suivante :

$$x \leq y \Rightarrow \lfloor x \rfloor \leq \lfloor y \rfloor \quad x \geq 0, y \geq 0$$

a) si $x = y$: trivial comme indiqué plus haut $x = y \Rightarrow \lfloor x \rfloor = \lfloor y \rfloor \Rightarrow \lfloor x \rfloor \leq \lfloor y \rfloor$

b) si $x < y$

$$x < y \text{ et } y < \lfloor y \rfloor + 1 \Rightarrow x < \lfloor y \rfloor + 1$$

comme $\lfloor x \rfloor \leq x$ alors

$$\lfloor x \rfloor < \lfloor y \rfloor + 1$$

étant donné que $\lfloor x \rfloor, \lfloor y \rfloor \in \mathbb{N}$ alors

$$\lfloor x \rfloor \leq \lfloor y \rfloor \text{ c.q.f.d.}$$

III.7.2 Qualité de la solution arrondie

La solution $\lfloor \tilde{X}^* \rfloor$ respecte donc toutes les contraintes de temps de travail. En revanche, cette solution ne respecte pas nécessairement les contraintes de couverture. Les périodes en sous-effectifs seront celles qui correspondent au non respect de ces contraintes. Nous donnons ci-dessous une borne supérieure au sous-effectif que peut induire cette solution.

a) *Pas de flux stockable*

Proposition :

Dans le cas où l'on a un ou plusieurs flux d'appels, la partie entière de la solution relaxée $\lfloor \tilde{X}^* \rfloor$ conduira à un sous-effectif par période qui sera au pire égal à $\sum_{i \in I} (\|A_{appels}(i)\|) - 1$

Avec

$\|A_{appels}(i)\|$ correspond au nombre d'activités d'appels du profil i .

Démonstration

Le sous-effectif que génère la partie entière de la solution relaxée $\lfloor \tilde{X}^* \rfloor$ s'exprime pour chaque période t comme la différence entre le total de la capacité de prise d'appels et le total de la charge d'appels à couvrir soit :

$$\text{sous-effective}(t) = \sum_{f \in F_{appels}} D_t^f - \sum_{i \in I, a \in A_{appels}} [\tilde{X}_t^a(i)] - [\tilde{X}_t^{a+1}(i)]$$

Nous devons démontrer donc que $\text{sous-effective}(t) \leq \sum_{i \in I} (\|A_{appels}(i)\|) - 1 \quad \forall t = \overline{T_0, T_{end}}$

Dans le cas d'un ou plusieurs flux d'appels, les contraintes de couverture s'expriment comme suit :

$$N_t^{appels}(i) \geq \sum_{f \in F_{appels}} p_t^f(i) \cdot D_t^f \quad \forall t = \overline{T_0, T_{end}}, \forall i \in I \quad \text{III-14}$$

En sommant III-14 sur I on obtient

$$\sum_{i \in I} N_t^{appels}(i) \geq \sum_{i \in I, f \in F_{appels}} p_t^f(i) \cdot D_t^f \quad \forall t = \overline{T_0, T_{end}}$$

Etant donné que $\sum_{i \in I} p_t^f(i) = 1 \quad \forall t = \overline{T_0, T_{end}}, \forall f \in F_{appels}$ alors

$$\sum_{i \in I} N_t^{appels}(i) \geq \sum_{f \in F_{appels}} D_t^f \quad \forall t = \overline{T_0, T_{end}}$$

Soit

$$\sum_{i \in I, a \in A_{appels}} \tilde{X}_t^a(i) - \tilde{X}_t^{a+1}(i) \geq \sum_{f \in F_{appels}} D_t^f \quad \forall t = \overline{T_0, T_{end}} \quad \text{III-15}$$

Par ailleurs, nous avons par définition :

$$[\tilde{X}_t^a(i)] \leq \tilde{X}_t^a(i) < [\tilde{X}_t^a(i)] + 1 \quad \forall a \in A_{appels}(i), i \in I \text{ donc}$$

$$\begin{cases} \tilde{X}_t^a(i) < [\tilde{X}_t^a(i)] + 1 \\ -\tilde{X}_t^{a+1}(i) \leq -[\tilde{X}_t^{a+1}(i)] \end{cases} \quad \forall a \in A_{appels}, i \in I$$

En sommant on trouve

$$\sum_{i \in I, a \in A_{appels}(i)} \tilde{X}_t^a(i) - \tilde{X}_t^{a+1}(i) < \sum_{i \in I, a \in A_{appels}(i)} ([\tilde{X}_t^a(i)] - [\tilde{X}_t^{a+1}(i)] + 1)$$

D'après III-15 on déduit

$$\sum_{f \in F_{appels}} D_t^f < \sum_{i \in I, a \in A_{appels}(i)} \left(\lfloor \tilde{X}_t^a(i) \rfloor - \lfloor \tilde{X}_t^{a+1}(i) \rfloor + 1 \right)$$

$$\sum_{f \in F_{appels}} D_t^f - \sum_{i \in I, a \in A_{appels}(i)} \left(\lfloor \tilde{X}_t^a(i) \rfloor - \lfloor \tilde{X}_t^{a+1}(i) \rfloor \right) < \sum_{i \in I, a \in A_{appels}(i)} 1 = \sum_{i \in I} \left(\|A_{appels}(i)\| \right)$$

Etant donné que $\sum_{f \in F_{appels}} D_t^f \in \mathbb{N}$ (car il correspond à un nombre de conseillers) alors

$$\sum_{f \in F_{appels}} D_t^f - \sum_{i \in I, a \in A_{appels}(i)} \left(\lfloor \tilde{X}_t^a(i) \rfloor - \lfloor \tilde{X}_t^{a+1}(i) \rfloor \right) \leq \sum_{i \in I} \left(\|A_{appels}(i)\| \right) - 1$$

$$\text{sous-effectif}(t) \leq \sum_{i \in I} \left(\|A_{appels}(i)\| \right) - 1 \quad \mathbf{c.q.f.d.}$$

Notons que cette borne ne dépend pas de la charge du centre d'appels. Ainsi, plus le centre d'appels est grand plus l'impact d'une solution arrondie par rapport à une solution exacte se confond avec les éléments d'incertitude lié à l'environnement du centre d'appels. Ainsi, si l'on considère par exemple un centre d'appels avec deux profils de conseillers : temps plein et temps partiel Les conseillers temps plein prennent une pause déjeuner $\|A_{appels}(\text{temps plein})\| = 2$ et les conseillers temps partiel ne prennent pas de pause $\|A_{appels}(\text{temps partiel})\| = 1$. Le sous-effectif généré par la partie entière de la solution relaxée sera donc de l'équivalent de 2 conseillers par période. Pour un centre d'appels de taille moyenne (quelques centaines de conseillers) l'écart représente moins de 1% de l'effectif à recruter ce qui est très négligeable compte tenu de l'incertitude sur les données liées à l'horizon temporel du dimensionnement (des erreurs de prévisions de l'ordre de 10% sont très communes dans la pratique) ou sur les taux d'absentéisme.

b) Cas avec flux stockables

Les contraintes de traitement des appels sont plus restrictives que celles des flux stockables. En effet, dans nos modèles, nous avons considéré une contrainte de couverture par période pour les appels. S'agissant des flux stockables qui ont un objectif de traitement sur la journée, nous avons considéré que leur traitement pouvait s'étendre sur plusieurs périodes. De ce fait, les sous-effectif induit par une solution arrondie pour un flux stockable est **au pire** égal au sous-effectif que l'on aurait s'il s'agissait d'un flux d'appels.

Ainsi, la borne supérieure précédente calculée pour les flux d'appels peut être généralisée aux flux stockables. Ces flux correspondent aux activités hors pause (appels et flux stockables)

Théorème :

Dans le cas où l'on a un ou plusieurs flux d'appels ou flux stockables, la partie entière de la solution relaxée $\lfloor \tilde{X}^* \rfloor$ conduira à un sous-effectif par période qui sera au pire égal à

$$\sum_{i \in I} (\|A_{hors-pause}(i)\|) - 1$$

Avec

$\|A_{hors-pause}(i)\|$ est le nombre d'activités du profil i qui ne correspondent pas à des pauses (appels et flux stockables)

III.8 Conclusion

La première décision en matière de gestion des ressources humaines dans un centre d'appels est celle du dimensionnement. Le modèle d'optimisation le plus largement utilisé pour répondre à cette problématique dans la littérature est le *shift scheduling*.

Pour répondre à cette problématique de dimensionnement, nous avons présenté dans ce chapitre le paradigme de chaîne d'activités. Ce paradigme nous a permis de mettre en œuvre des développements qui ont abouti à quelques résultats intéressants.

a) *Le paradigme de chaîne d'activités : une représentation non mathématique*

Le premier intérêt du paradigme de chaîne d'activités est qu'il permet de représenter à partir de notions simples (activités, profils, dates de début d'activités, etc.) la grande diversité des environnements et des contraintes de gestion des ressources humaines dans un centre d'appels. De plus, ce paradigme ne nécessite pas de recourir à un formalisme mathématique peu accessible à des non initiés.

b) *Le paradigme se décline en un modèle implicite, générique et unificateur*

Nous avons montré dans ce chapitre comment déduire de notre paradigme le PLNE qui permet de résoudre le problème *shift scheduling*. Le modèle que nous proposons est implicite dans la continuité de ceux que nous avons présentés dans l'état de l'art. Il présente l'avantage de ne pas imposer une énumération explicite de toutes les combinaisons d'horaires possibles.

De plus, la généralité du paradigme de chaîne d'activités se retrouve dans notre modèle. Il ne présente donc pas les mêmes limitations que les précédents modèles de la revue bibliographique concernant le nombre de pauses et de profils. Toutes les spécificités en terme de contraintes de temps de travail sont prises en compte par notre PLNE dès lors qu'elles s'inscrivent dans le cadre du paradigme de chaîne d'activités.

Enfin, notre modèle élargi le périmètre du problème de *shift scheduling* généralement étudié. En effet, en plus des activités d'appels et de pause, nous intégrons les notions de multi-compétence, de flux stockable et de *call blending*. Notons que ces aspects sont pris en compte dans notre modèle sans nécessiter de modifications des hypothèses définissant le cadre du paradigme de chaînes d'activités.

c) La totale unimodularité et la difficulté du problème de shift scheduling

Nous avons montré dans ce chapitre que la structure de la matrice des contraintes de temps de travail était totalement unimodulaire. Ceci nous a permis de mettre en évidence deux résultats :

- Les contraintes de couverture induisent la NP-difficulté du problème de *shift-scheduling*.
- Nous avons pu identifier des conditions particulières pour lesquelles le problèmes de *shift-scheduling* peut être résolu en un temps polynomial.

d) Estimation d'une borne supérieure pour la solution arrondie de notre modèle

Le modèle que nous proposons est un PLNE. Dans certains cas, il peut être intéressant d'obtenir des solutions approchées telles que la résolution du programme linéaire en relaxant les contraintes d'intégrité. Dans ce chapitre, nous avons calculé une borne supérieure de l'écart que donne la solution arrondie d'un tel problème par rapport à la solution exacte du PLNE.

Chapitre IV

PLANIFICATION MAXIMISANT LA QUALITÉ DE SERVICE

Dans le chapitre précédent, la contrainte de qualité de service était un paramètre d'entrée pour constituer les contraintes de couverture. Elle était prise en compte en amont du problème de construction de vacations. C'est l'approche classique utilisée dans les contributions scientifiques concernant ce problème. Dans le présent chapitre, nous nous situons dans la phase de planification. Les conseillers de clientèle sont déjà recrutés (lors de la précédente phase de dimensionnement) et l'objectif est de les affecter à des vacations de façon à maximiser la qualité de service. Il ne s'agit plus d'évaluer la qualité de service séparément du problème de construction de vacations mais de l'intégrer dans le modèle de planification.

Après un bref état de l'art, nous présentons dans ce chapitre notre modèle issu du paradigme de chaîne d'activités pour formuler et résoudre ce problème. Nous terminons par une application numérique sur un exemple de test.

IV.1 Position du problème

Dans ce chapitre, nous nous intéressons au problème de planification horaire décrit dans la Figure III-1. Nous considérons que les conseillers sont déjà recrutés et que leurs jours d'absence et de présence sont planifiés. Les questions de minimisation du coût salarial ainsi que celle de la planification journalière sont donc traitées en amont.

L'objectif de la planification horaire est de garantir la meilleure qualité de service aux clients compte tenu des effectifs présents. A ce niveau, deux décisions doivent être prises :

- Construire les vacances à proposer aux conseillers
- Déterminer le nombre de conseillers nécessaires pour chaque vacation

Le résultat de ces deux décisions servira de base pour construire les emplois du temps individuels des conseillers.

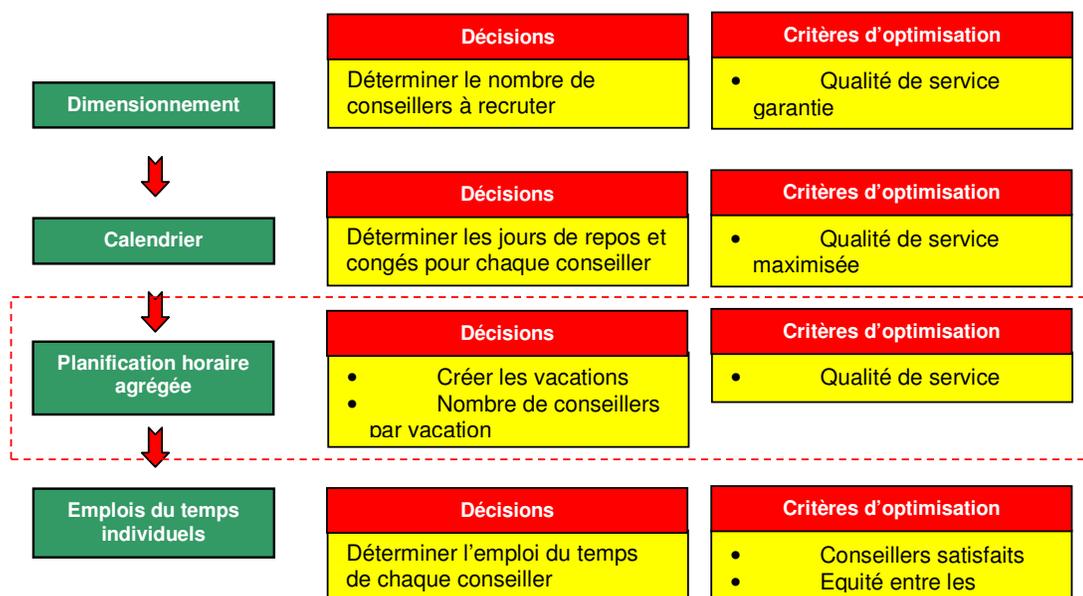


Figure IV-1 : Processus de décision englobant le problème de planification avec maximisation de la qualité de service

Remarque : nous ne nous intéressons pas dans ce qui suit à la problématique d'infaisabilité ou de sous-optimalité induite par la décomposition entre planification journalière et planification horaire. Ceci se gère dans la pratique, par exemple, en planifiant chaque journée séparément à partir du nombre de conseillers présents (qui ne sont pas en

congé ou en repos) ou alors en se donnant des règles empiriques permettant de surévaluer le nombre de conseillers lors du dimensionnement. Cette problématique est préalablement traitée par l'entreprise lorsqu'elle décide de la manière de décomposer son processus de décision de gestion des ressources humaines. Nous ne reviendrons donc pas sur cet aspect que nous avons évoqué dans le chapitre 2 sur l'organisation des décisions.

Dans ce qui suit, nous proposons de résoudre ce problème de planification agrégée par une adaptation du modèle de *shift scheduling* issu du paradigme de chaînes d'activités en y intégrant un objectif de qualité de service globale sur la journée. Avant de présenter ce modèle nous nous attarderons au préalable sur les difficultés induites par l'intégration d'objectifs de qualité de service dans le *shift scheduling* et son traitement dans la littérature.

IV.2 Prise en compte de la qualité de service

IV.2.1 Qualité de service à la journée plutôt que par période

Dans la vision classique du *shift scheduling*, on prend comme hypothèse que la qualité de service est imposée par période (par exemple, 80% des appels dans la période d'une demi-heure traités). Dans un centre d'appels et plus généralement dans les services, cette vision est très contraignante. En réalité, il n'est pas nécessaire de garantir une qualité de service constante à chaque période de la journée. On peut se permettre en pratique de compenser les "mauvaises périodes" par de "bonnes périodes" de la journée (Koole et van_der_Sluis 2003). Cette vision plus souple et plus réaliste permet in fine une meilleure adéquation entre la charge et la capacité. Dans le problème que nous abordons, nous considérons donc un objectif de qualité de service à maximiser globalement sur la journée et non plus par période. Ceci conduit de fait à prendre en compte explicitement la qualité de service dans le modèle puisqu'il n'est plus possible de déterminer le nombre de conseillers nécessaires par période ou le volume de travail à fournir sur la journée.

IV.2.2 Impact de la non-linéarité de la qualité de service sur le shift scheduling

La qualité de service (temps moyen d'attente, proportion d'appels répondus, etc.) est le plus souvent modélisées par des files d'attente qui ajoutent une non-linéarité au critère à optimiser. L'intégration de cette non-linéarité au *shift scheduling* implique une difficulté supplémentaire à un problème déjà combinatoire. Cette difficulté explique, pour une part, le fait que peu de contributions dans la littérature s'intéressent au *shift scheduling* maximisant un critère de qualité de service non linéaire.

Une approche alternative de simplification consiste à prendre comme critère de qualité de service la différence entre capacité et charge. C'est une évaluation linéaire mais qui ne prend pas en compte le caractère stochastique du fonctionnement d'un centre d'appels. On parle dans ce cas d'approche déterministe par opposition à une approche stochastique. Elle présente l'avantage de maintenir une structure linéaire du problème de *shift scheduling* et de permettre une résolution plus simple. Cependant, cette approche ne rend pas compte du gain décroissant sur la qualité de service d'une augmentation de capacité. Autrement dit, l'ajout d'un 3^e conseiller apporte moins en terme de qualité de service que le 2nd.

L'approche que nous proposons est encore différente. Elle procède par linéarisations. Elle permet d'intégrer un objectif de qualité de service non linéaire dans le *shift scheduling* tout en préservant la linéarité du problème.

IV.2.3 Etat de l'art

Dans le chapitre 3, la qualité de service est calculée en amont de la résolution du problème de *shift scheduling*. Cette séparation entre évaluation de la qualité de service et le *shift scheduling* n'est plus possible dès lors que l'on considère que les périodes de planification ne sont plus indépendantes entre elles.

Nous pouvons identifier dans la littérature deux cas de figure où l'indépendance des périodes de planification ne peut pas être garantie :

Couplage des périodes par la fonction objectif : c'est le cas que nous avons décrit plus haut où la finalité est de maximiser la qualité de service globalement sur toute la journée par opposition à une qualité de service par période. Chaque période influence la qualité de service globale et elles ne peuvent donc être traitées séparément.

Couplage lié à l'influence mutuelle entre périodes : dans certaines situations, les périodes de planification sont dépendantes en raison du caractère transitoire des phénomènes sous-jacents. C'est le cas, par exemple, lorsque les périodes de planification sont trop courtes par rapport au temps de service. Les situations où les rappels clients sont conséquents conduisent également à lier les périodes entre elles. En effet, dans ce cas, la qualité de service d'une période dépend du nombre de rappels (et donc de la qualité de service) issus de la période qui la précède.

Dans ce qui suit nous présentons quelques contributions présentant ces deux aspects. Ce sont des travaux de modélisation et de résolution du problème de *shift scheduling* avec prise en compte explicite de la qualité de service. Les modèles concernent des problématiques de dimensionnement (déterminer le nombre de conseiller à recruter) ou de planification (organiser le temps de travail des conseillers pour maximiser la qualité de service)

IV.2.3.1 Couplage par la fonction objectif : qualité de service globale à la journée

(Koole et van_der_Sluis 2003) abordent la problématique d'optimisation du *shift scheduling* dans un centre d'appels en prenant en compte une qualité de service globale sur la journée.

Le modèle proposé repose sur une représentation explicite des vacations dans un centre d'appels. Toutes les vacations sont de durées identiques. Ils supposent également qu'il n'y a pas de pauses. Ce modèle permet de résoudre deux problèmes. Le premier est un problème de dimensionnement visant à minimiser les effectifs compte tenu d'une contrainte de qualité de service globale sur la journée. Le second problème traité est un problème de planification visant à minimiser cette qualité de service globale à effectif donné.

L'approche utilisée pour résoudre le problème de planification est une méthode de recherche locale. Compte tenu des hypothèses prises et d'une propriété du problème appelée multimodularité, (Koole et van_der_Sluis 2003) démontrent qu'une méthode de recherche locale aboutit à la solution exacte du problème. Pour plus de détails concernant la multimodularité consulter (Hajek 1985; Altman, Gaujal et al. 2000).

La résolution du problème de dimensionnement se fait par dichotomie. Le problème de planification cité plus haut est résolu à chaque itération en faisant varier le nombre d'effectifs. L'algorithme converge vers la solution optimale de N conseillers lorsque la contrainte de qualité de service est respectée mais qu'elle n'est pas respectée pour $N-1$ conseillers.

IV.2.3.2 Couplage par influence mutuelle entre périodes

Les approches en régime transitoire sont pertinentes lorsque l'évolution du système ne peut être étudiée en considérant les périodes de planification indépendamment les unes des autres. C'est le cas lorsque les temps de service sont relativement longs par rapport à la durée d'une période de planification. Prenons par exemple le cas d'un problème de planification des patrouilles de police. La durée de la période de planification est d'une demi-heure et la durée moyenne d'une patrouille est de plusieurs heures (temps de service). Dans ce cas, il est évident qu'il n'est pas possible de considérer que les phénomènes sont indépendants et stationnaires au sein de chaque période de planification. Il est nécessaire de tenir compte de l'effet d'une période sur les autres. L'approche en régime transitoire est également appropriée lorsqu'il s'agit de modéliser les effets de surcharge (*rush hour*). En effet, en régime stationnaire, les évaluations de performance des files d'attente (temps moyen d'attente) sont indéterminées lorsque la charge est temporairement supérieure à la capacité de service et ne peuvent donc être traitées avec des approches classiques. Toutefois, il est à noter que ce cas de figure est assez rare dans la pratique. Du moins, ce type d'approche ne devrait pas être utilisé pour des problèmes de planification ou de dimensionnement dont l'objectif est de garantir ou maximiser la qualité de service. En effet, avoir des périodes où la charge est supérieure à la capacité de service peut conduire à des instabilités du système qui rendent la solution très peu robuste. La qualité de service obtenue n'est ainsi pas maîtrisée.

Compte tenu de la complexité des phénomènes transitoires, ils se prêtent difficilement à des modélisations analytiques. C'est la raison pour laquelle l'essentiel des travaux prenant en compte ces phénomènes dans le *shift scheduling* ont recours à des méthodes non analytiques telles que les méta-heuristiques ou la simulation.

Approche méta-heuristique (algorithmes génétiques)

(Ingolfsson, Haque et al. 2002) s'est intéressé au problème de planification de patrouilles de police. Son approche permet d'intégrer la qualité de service en tant que contrainte à respecter ou en tant qu'objectif à optimiser. L'approche permet également de prendre en compte une qualité de service par période ou globale sur la journée.

La technique de résolution proposée par (Ingolfsson, Haque et al. 2002) est une méthode heuristique qui repose sur deux composants :

- Un composant d'évaluation de la qualité de service (*service level evaluator*) : il permet d'évaluer la qualité de service à chaque période sur la base d'une méthode numérique
- Un composant de génération de planification (*schedule generator*) : il repose sur une technique d'algorithme génétique qui sélectionne les meilleures planifications.

Une amélioration de cette méthode est présentée dans (Ingolfsson, Cabral et al. 2002). Il s'agit d'une approche heuristique reposant sur la programmation linéaire qui permet un temps de résolution plus rapide.

Approche par simulation

Un autre type d'approche de résolution consiste à intégrer l'aspect transitoire de l'évaluation de qualité de service par la simulation événementielle. La résolution consiste à effectuer des itérations entre programme linéaire et simulation. Cette approche a été formalisée dans (Henderson et Mason 1998). Nous pouvons trouver une illustration dans (Atlason, Epelman et al. 2004; Atlason, Epelman et al. 2005) qui développent une approche de plan de coupes associées à la simulation. Dans (Atlason, Epelman et al. 2008) cette approche est généralisée pour prendre en compte des évaluations de qualité de service non convexe (en forme de S).

IV.3 Modélisation

IV.3.1 Hypothèses

Pour la construction du modèle nous reprenons le paradigme de chaîne d'activités avec les mêmes hypothèses décrites dans le chapitre 3.

- La durée de travail est inférieure à 24h.
- Les conseillers appartiennent à des profils. A chaque profil est associé un ensemble de contraintes de temps de travail identiques et un enchaînement d'activités préalablement défini.
- On suppose dans ce cas qu'il y a un seul flux d'appels à traiter (pas de multi-flux, ni de flux stockables)

Le modèle que nous présentons est similaire à celui décrit dans le précédent chapitre en particulier en ce qui concerne les contraintes de temps de travail. Les évolutions que nous apportons concernent deux aspects :

- **Ajout de contraintes de limitation des effectifs** : la capacité de service est limitée par le nombre de conseillers recrutés lors de la phase antérieure de dimensionnement.
- **Modification de la fonction objectif** : nous n'avons plus de contraintes de couverture. Il est nécessaire d'introduire dans le modèle une formulation de la qualité de service à optimiser.

IV.3.2 Contraintes

Nous utilisons le paradigme de chaîne d'activités développé dans le chapitre précédent pour modéliser le problème. Nous gardons les mêmes variables de décision et les contraintes de temps de travail s'expriment de la même manière. Soit :

Contraintes d'intégrité

$$X_t^a(i) \in \mathbb{N} \quad \forall i \in I, \forall a \in A(i), \forall t = \overline{T_0, T_{end}} \quad \text{IV-1}$$

Contraintes de temps maximum et minimum de travail

$$\begin{cases} X_{t-l_{a,a'}}^a(i) \geq X_t^{a'}(i) & l_{a,a'}(i) \leq t \leq T_{end} \\ X_t^{a'}(i) \geq X_{t-L_{a,a'}}^a(i) & L_{a,a'}(i) \leq t \leq T_{end} \end{cases} \quad \text{IV-2}$$

Contraintes de fenêtres de temps

$$\begin{cases} X_{T_{\min}^a(i)-1}^a(i) = 0 \text{ avec } T_{\min}^a(i) \neq 0 \\ X_{T_{\max}^a(i)}^a(i) = X_{T_{end}}^a(i) = N(i) \end{cases} \quad \text{IV-3}$$

Contraintes de successions d'activités

$$X_t^a(i) \geq X_t^{a+1}(i) \quad \forall i \in I, \forall a \in A(i) - \{a_{end}\}, \forall t = \overline{T_0, T_{end}} \quad \text{IV-4}$$

Variables de cumul croissantes

$$X_{t+1}^a(i) \geq X_t^a(i) \quad \forall i \in I, \forall a \in A_i, \forall t = \overline{T_0, T_{end} - 1} \quad \text{IV-5}$$

Ajout des contraintes de limitation des effectifs

Si Y_i^{\max} correspond au nombre de conseillers disponibles pour le profil i , la contrainte de limitation du nombre de conseillers par profil peut s'exprimer comme suit :

$$X_{T_{end}}^{a_{end}}(i) = Y_i \leq Y_i^{\max} \quad \forall i \in I \quad \text{IV-6}$$

IV.3.3 Fonction objectif – optimisation de la qualité de service

L'indicateur de qualité de service que nous prenons comme objectif est la minimisation du nombre de clients dits déconnectés. C'est à dire les clients à qui on demande de rappeler ultérieurement. Soit dec_t , le nombre d'appels déconnectés durant la période t . L'objectif s'exprime donc comme suit :

$$\min Z = \sum_{t=T_0}^{T_{end}} dec_t \quad \text{IV-7}$$

Remarque :

Dans le cas où l'on souhaiterait maximiser la qualité de service par période et non plus à la journée, l'objectif serait alors d'optimiser la plus mauvaise période. C'est-à-dire minimiser le taux de déconnexion maximum de la journée.

Notons dec^{max} la variable de décision correspondant au taux de déconnexion le plus élevé de la journée. On doit introduire les contraintes suivantes

$$dec^{max} \geq dec_t \quad \forall t = \overline{T_0, T_{end}} \quad \text{IV-8}$$

La fonction à optimiser deviendrait alors

$$\min Z = dec^{max} \quad \text{IV-9}$$

IV.3.3.1 Deux approches de modélisation de la qualité de service

Le centre d'appels est un système où les phénomènes sont fondamentalement aléatoires. En effet, les appels arrivent selon un processus stochastique (l'hypothèse la plus courante est qu'il suit une loi de Poisson) et les temps de réponses ne sont pas constants. Dans ce qui suit nous nous intéresserons à deux manières d'aborder la modélisation de la qualité de service : une approche stochastique basée sur un modèle de file d'attente permettant de prendre en compte les phénomènes aléatoires et une approche déterministe qui simplifie la première en supposant que l'aléa n'intervient pas.

a) Approche déterministe

Si on ne prend pas en compte l'aléa, le nombre d'appels déconnectés par période est la différence entre la charge d'appels et la capacité de service (c'est à dire le nombre d'appels qu'il est possible de servir par période). La Figure IV-2 représente graphiquement le nombre d'appels déconnectés sur une journée en fonction de l'évolution de la charge d'appels et de la capacité de service. Notons que cette approche revient à minimiser le sous-effectif.

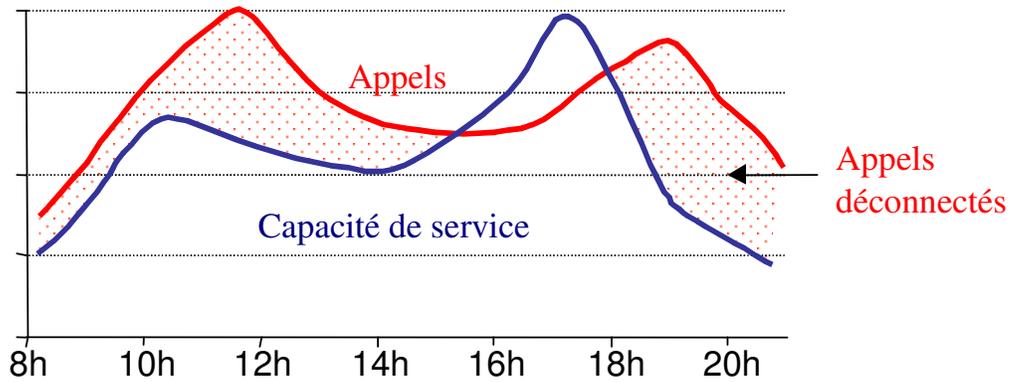


Figure IV-2: Illustration de la déconnexion sur une journée

Ainsi, au cours d'une période t le nombre d'appels déconnectés dec_t est donné par :

$$dec_t = \max(charge_t - \mu \cdot N_t^{appels}, 0)$$

Rappelons que :

$$N_t^{appels} = \sum_{i \in I, a \in A_{appels}(i)} N_t^a(i)$$

$A_{appels}(i)$: ensemble des activités des conseillers qui correspondent à de la prise d'appels pour le profil i

N_t^{appels} : nombre total de conseillers en prise d'appels à la période t

$charge_t$: nombre d'appels qui arrivent durant la période t .

μ : taux de service, i.e. nombre d'appels servis par un conseiller pendant une période

Pour obtenir une évaluation linéaire du nombre d'appels déconnectés nous pouvons introduire les contraintes suivantes :

$$dec_t \geq charge_t - \mu \sum_{i \in I, a \in A_{appels}} N_t^a(i) \quad \forall t = \overline{T_0, T_{end}}$$

La prise en compte de la qualité de service de manière déterministe nous conduit donc à ajouter au modèle $(T_{end} - T_0 + 1)$ variables de décision dec_t et autant de contraintes.

b) Approche stochastique

Afin de tenir compte des aléas, nous utiliserons une file d'attente pour modéliser le centre d'appels. Afin d'obtenir des évaluations continues nous utiliserons une file M/M/1/K pour approcher une file M/M/C/K. K étant la taille de la file d'attente et C le nombre de conseillers. Le taux de service de la file M/M/1/K équivalente sera $C \cdot \mu$.

La proportion d'appels déconnectés est donnée par la probabilité de blocage de la file d'attente. Elle s'exprime de la manière suivante :

$$P_t^{blocage}(N_t^{appels}) = \frac{\left[\left(\frac{N_t^{appels} \cdot \mu}{charge_t} \right) - 1 \right]}{\left[\left(\frac{N_t^{appels} \cdot \mu}{charge_t} \right)^{K+1} - 1 \right]}$$

Le nombre d'appels déconnectés à la période t est donné par :

$$dec_t = charge_t \cdot P_t^{blocage}(N_t^{appels})$$

Dans la Figure IV-3 est représentée la probabilité de blocage, c'est à dire la proportion d'appels déconnectés, en fonction de la capacité de service ($N_t^{appels} \cdot \mu$). Cette probabilité est égale à 1 lorsque le nombre de conseillers est nul et elle tend vers zéro à mesure que ce nombre augmente. La proportion d'appels déconnectés est également représentée pour le cas déterministe. Contrairement au cas stochastique, cette proportion s'annule si la capacité de service correspond à la charge d'appels. L'approche déterministe sous-estime systématiquement la proportion d'appels déconnectés par rapport à l'approche stochastique car elle ne prend pas en compte la disparité du temps de service des conseillers ni l'arrivée irrégulière des clients.

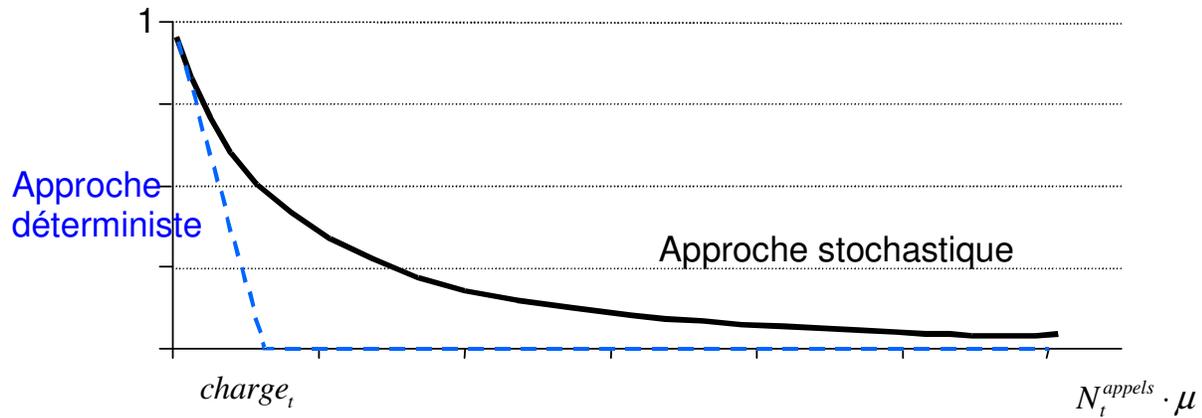


Figure IV-3: Probabilité de blocage dans le cas déterministe et stochastique

L'évaluation de la probabilité est une fonction non linéaire du nombre de conseillers. Afin de pouvoir utiliser cette évaluation dans une formulation en programme linéaire, il est nécessaire de linéariser cette fonction.

Le principe de la linéarisation consiste à approcher une fonction non-linéaire en introduisant un ensemble de contraintes linéaires. La linéarisation de la proportion de déconnectés en fonction du nombre de conseillers en prise d'appels N_t^{Appels} est illustrée dans la Figure IV-4.

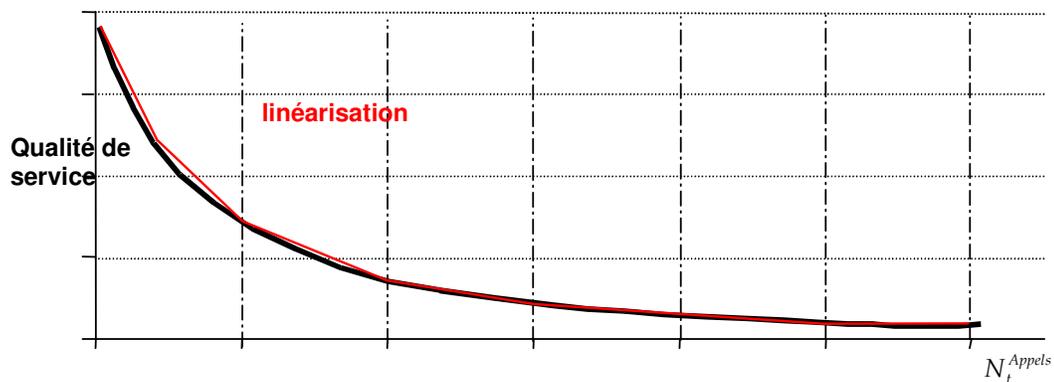


Figure IV-4: Linéarisation de l'évaluation de la qualité de service

Comme notre fonction à linéariser est convexe, la linéarisation évalue de manière légèrement pessimiste le nombre de déconnectés. L'erreur d'approximation ne dépend ici

que du choix du nombre et des points de linéarisation. Comme N_t^{Appels} est entier, il suffit de prendre comme points de linéarisation toutes ses valeurs entières possibles pour que l'approximation soit exacte.

Pour que l'utilisation de ce principe de linéarisation soit possible dans un programme linéaire, il est suffisant que la fonction à linéariser soit convexe en fonction de N_t^{Appels} . La convexité de la probabilité de blocage d'une file M/M/1/K est démontrée dans (Nagarajan et Towsley 1992).

Soient N_0, N_1, \dots, N_k un ensemble de valeurs à partir desquelles seront effectuées la linéarisation de la fonction $P_t^{blocage}(N_t^{appels})$.

On suppose que $N_0 = 0$ et que $N_0 < N_1 < \dots < N_k$ et posons :

$$P_t^k = P_t^{blocage}(N_k)$$

Les contraintes introduites seront celles qui correspondent aux droites reliant les points (N_k, P_t^k) et (N_{k+1}, P_t^{k+1}) . L'ensemble de contraintes peut ainsi être exprimé de la manière suivante :

$$dec_t \geq charge_t \cdot \left[\frac{P_t^{k+1} - P_t^k}{N_{k+1} - N_k} \cdot N + \frac{P_t^k \cdot N_{k+1} - P_t^{k+1} \cdot N_k}{N_{k+1} - N_k} \right] \quad k = 0 \text{ à } K - 1$$

Plus le nombre de points est grand plus la linéarisation est proche de la fonction que l'on souhaite approcher.

Si l'on suppose qu'à chaque période le nombre de conseillers ne peut dépasser un maximum N^{max} alors il est possible d'effectuer une linéarisation qui évalue de manière exacte la fonction pour les valeurs entières. Pour ce faire, il suffit de prendre comme points de linéarisation les valeurs entières comprises entre 0 et N^{max} avec $N_k = k$. On obtient alors, pour chaque période, un ensemble de N^{max} contraintes exprimées de la manière suivante :

$$dec_t \geq charge_t \cdot \left[(P_t^{k+1} - P_t^k) \cdot N_t^{Appels} + (k+1) \cdot P_t^k - k \cdot P_t^{k+1} \right] \quad k = 0 \text{ à } N^{max} - 1$$

Ainsi, l'intégration de la qualité de service **de manière exacte** nécessite $(T_{end} - T_0 + 1)$ variables de décision dec_t et $N^{max} \cdot (T_{end} - T_0 + 1)$ contraintes.

Notons que cette formulation ne se limite pas à la fonction d'évaluation de la qualité de service que nous avons choisie (proportion de déconnectés). Elle peut être généralisée à d'autres fonctions qui vérifient la condition de convexité.

IV.3.4 Synthèse du modèle

Voici un récapitulatif de notre modélisation du problème de planification avec maximisation de qualité de service

IV.3.4.1 Contraintes de temps de travail

Elles sont identiques à celles du problème de dimensionnement présentées au chapitre 3 (contraintes IV-1 à IV-5)

IV.3.4.2 Contraintes de limitation des effectifs

Si Y_i^{\max} correspond au nombre de conseillers disponibles pour le profil i , la contrainte de limitation du nombre de conseillers par profil peut s'exprimer comme suit :

$$X_{T_{end}}^{a_{end}}(i) = Y_i \leq Y_i^{\max} \quad \forall i \in I$$

IV.3.4.3 Fonction objectif

$$\min Z = \sum_{t=T_0}^{T_{end}} dec_t$$

a) Approche déterministe

Ajouter les contraintes suivantes :

$$dec_t \geq charge_t - \mu \sum_{i \in I, i \in A_{appels}} N_t^a(i) \quad \forall t = \overline{T_0, T_{end}}$$

b) Approche stochastique

Ajouter les contraintes suivantes

$$dec_t \geq charge_t \cdot \left[(P_t^{k+1} - P_t^k) \cdot N_t^{Appels} + (k+1) \cdot P_t^k - k \cdot P_t^{k+1} \right] \quad k = 0 \text{ à } N^{\max} - 1$$

Remarque 1 : généralisation de l'approche stochastique

L'approche de linéarisation de la qualité de service peut être généralisée à d'autres modèles que la file M/M/1/K. Pour ce faire, il suffit que la fonction représentant la qualité de service soit convexe en N_t^{Appels} (concave en cas de maximisation). L'approche proposée est ainsi valide pour des files d'attente de type Erlang A en choisissant un critère respectant la contrainte de convexité tels que décrits dans (Koole 2004; Koole 2005). Pour une vue plus générale de l'utilisation des critères de qualité de service dans les problèmes de planification voir (Whitt 2007).

Remarque 2 :NP-difficulté du problème

Notons que pour les deux approches, déterministe et stochastique, la structure des contraintes introduites par la fonction objectif n'est pas de la forme

$$x \leq y$$

$$x = y$$

Nous pouvons donc suivre le même raisonnement que le chapitre précédent et conclure de la même manière à un lien entre la NP-difficulté du problème de planification et les contraintes que nous introduisons pour prendre en compte la qualité de service (quelle que soit l'approche, stochastique ou déterministe).

IV.4 Comparaison entre la vision déterministe et la vision stochastique

Dans ce qui précède, nous avons détaillé deux approches pour intégrer la qualité de service dans le problème de *shift-scheduling*: l'approche déterministe et l'approche stochastique.

Nous proposons dans ce qui suit de comparer ces deux approches et de mettre en perspective l'apport de l'approche stochastique.

L'objectif n'est pas d'étudier de manière exhaustive et générale chaque approche. En effet, ceci nous écarterait du cadre de contribution de cette thèse qui se focalise d'abord sur des aspects de représentation du problème de shift scheduling et ses formulations en modèles mathématiques génériques (le paradigme de chaînes d'activité et les PLNE qui en découlent). La problématique de résolution et son pendant d'études numériques détaillées sont donc hors de périmètre de cette thèse. En revanche, à des fins d'illustration, nous appliquons dans ce qui suit notre modélisation sur un exemple de centre d'appels.

a) Méthodologie utilisée

Nous utilisons notre modèle pour résoudre le problème de planification avec maximisation de la qualité de service. Nous faisons intervenir les deux approches décrites précédemment : déterministe et stochastique.

L'exemple que nous considérons est inspiré du cas réel d'un opérateur de téléphonie mobile.

Pour comparer l'intérêt des deux approches d'évaluation de la qualité de service, nous analysons l'impact de chaque méthode de planification sur la qualité de service obtenue.

Afin de balayer un large éventail de situations en terme d'adéquation entre charge et capacité nous effectuons des tests numériques pour différentes valeurs de la charge d'appels en maintenant les effectifs constants.

b) Description de l'exemple de test

Le centre d'appels que nous étudions est opérationnel entre 8h et 21h. Il emploie deux types de conseillers regroupés en deux profils :

Profil temps plein

- Nombre de conseillers : 200
- Temps minimum de travail : 4h

- Temps maximum de travail : 8h
- Pause déjeuner d'une heure entre 11h30 et 13h30

Profil temps partiel

- Nombre de conseillers : 40
- Temps minimum de travail : 2h
- Temps maximum de travail : 5h
- Pause déjeuner d'une demi-heure

Le critère de qualité de service mesuré est la **proportion d'appels déconnectés sur la journée.**

Pour chaque test, nous faisons varier proportionnellement la charge d'appels initiale. La charge d'appels de chaque période est multipliée par le même facteur. Ce facteur multiplicatif varie entre 0.4 et 2.

c) Résultats des tests

Nous mesurons deux évaluations :

- La proportion d'appels déconnectés obtenue sur la journée en planifiant avec l'approche déterministe d'évaluation de la qualité de service.
- La réduction du nombre de déconnectés que permet l'utilisation de l'approche stochastique par rapport à l'approche déterministe d'évaluation de la qualité de service.

Les résultats des tests sont illustrés dans la Figure IV-5. Deux observations peuvent être faites :

- La proportion d'appels déconnectés est croissante en fonction de la charge d'appels. Ceci est une constatation évidente du fait que le nombre de conseillers et donc la capacité de service globale reste inchangé.
- L'amélioration apportée par l'approche stochastique est d'autant plus significative en proportion que la charge d'appels est réduite. Ceci s'explique par le fait que l'approche déterministe "se contente" seulement de minimiser le sous-effectif. Cette approche n'évalue pas, par définition, les conséquences d'une capacité de service supérieure à la charge d'appels sur la qualité de service. Il est donc naturel que l'approche déterministe soit moins performante que l'approche stochastique

lorsque le sur-effectif est important. Ce qui revient, dans notre exemple, au cas où la charge d'appels est faible par rapport à la capacité de service.

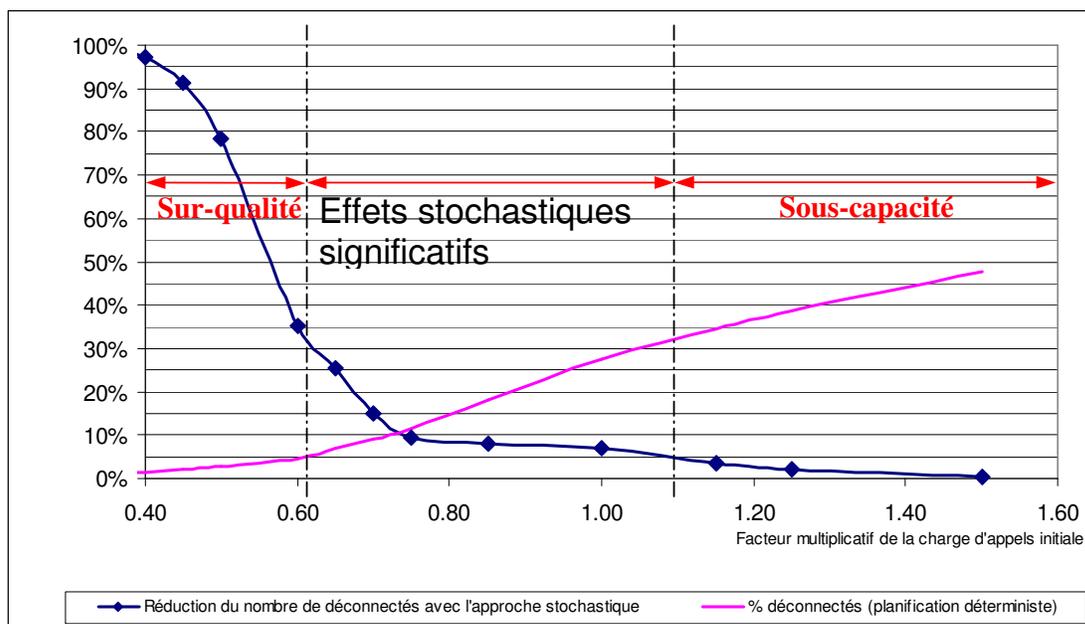


Figure IV-5: Comparaison de l'impact sur la planification d'une approche déterministe et stochastique d'évaluation de la qualité de service

d) Analyses et conclusions

En partant de ces deux observations, nous pouvons apprécier l'intérêt de l'approche stochastique par rapport à l'approche déterministe en fonction de la situation du centre d'appels. Nous pouvons, en effet, distinguer trois cas de figure (mis en évidence dans la Figure IV-5) pour lesquels l'apport d'une approche stochastique n'est pas identique.

Situation de sur-qualité (surcapacité) : (% déconnectés < 5%)

Dans ce cas, la qualité de service est déjà importante. Dans l'exemple, nous supposons qu'une proportion d'appels déconnectés inférieure à 5% représente une qualité de service très significative (ce seuil à 5% est à l'appréciation de l'entreprise, le même raisonnement peut s'appliquer à d'autres valeurs : 1%, 10%, etc.). Aussi, bien qu'en proportion l'approche stochastique soit très avantageuse, son intérêt est très limité car la marge d'amélioration est très réduite.

Situation de sous-capacité (%amélioration de l'approche stochastique <5%)

Dans ce cas le nombre de déconnectés est important mais l'amélioration apportée par l'approche stochastique n'est pas significative. En effet, dans notre cas nous considérons qu'une amélioration inférieure à 5% n'est pas significative. De la même manière que précédemment, ce seuil est fixé pour l'exemple en question, il n'est donné qu'à titre indicatif. Cette inefficacité de l'approche stochastique est liée au fait que le centre d'appels est en situation de sous-capacité au point qu'il ne peut respecter la charge d'appels même en négligeant l'effet des aléas. Ce cas de figure est généralement lié à des problèmes de dimensionnement où l'effectif disponible n'est pas suffisant. L'utilisation d'une approche stochastique pour la planification présente donc peu d'intérêt par rapport à une approche déterministe.

Situation avec effets stochastiques significatifs (%déconnectés >5% et %amélioration de l'approche stochastique >5%)

Dans ce cas de figure la proportion d'appels déconnectés du fait d'une planification par approche déterministe est significative. Dans le même temps, l'approche stochastique apporte une amélioration sensible par rapport à l'approche déterministe. Les deux évaluations sont supérieures au seuil de 5% que nous avons fixés. C'est bien dans cette situation qu'une approche stochastique présente tout son intérêt par rapport à une approche déterministe.

IV.5 Conclusion

Nous nous sommes intéressés dans ce chapitre au problème de planification d'un centre d'appels ayant pour objectif la maximisation de la qualité de service.

Après un bref état de l'art, nous avons proposé, sur la base du paradigme de chaîne d'activités, un modèle permettant de formuler le problème. Ce modèle permet de prendre en compte la qualité de service selon deux approches: une approche déterministe qui correspond aux méthodes utilisées habituellement et une approche stochastique qui permet de prendre en compte les phénomènes aléatoires dans la planification.

L'approche stochastique, par définition plus précise que l'approche déterministe, a été utilisée dans un exemple représentatif de la réalité. Nous avons montré que, sous certaines conditions d'adéquation entre charge et capacité, cette approche pouvait représenter de réels gains en terme de qualité de service.

CONCLUSIONS ET PERSPECTIVES

Conclusions

Cette thèse aborde la problématique de gestion des ressources humaines (*workforce management*) dans un centre d'appels. C'est un élément structurant dans un centre d'appels où la masse salariale représente l'essentiel des coûts. D'un point de vue industriel, nous avons essayé d'apporter des éclairages concernant trois types de décisions fondamentales :

- **Les décisions de dimensionnement** : combien faut-il recruter de conseillers ? Ce problème est traité dans le chapitre 3.
- **Les décisions de planification** : comment organiser le temps de travail des conseillers une fois ceux-ci recrutés ? Ce problème est traité dans le chapitre 4.
- **Les décisions "d'organisation des décisions"** : comment organiser, décomposer et hiérarchiser les décisions ? Ce sont en quelque sorte des "méta-décisions" préalables à toutes les autres en matière de gestion des ressources humaines. Cet aspect est généralement peu questionné dans la littérature. Nous avons tenté d'esquisser quelques analyses sur le sujet dans le chapitre 2.

La démarche de recherche que nous avons utilisée pour aborder ces questions a consisté à nous focaliser sur le problème de *shift-scheduling* (problème de construction de vacations). C'est un problème qui est au cœur de très nombreux travaux de recherche relevant de la gestion des ressources humaines. Compte tenu de l'état de l'art et de notre propre expérience opérationnelle dans le domaine des centres d'appels, nous avons identifié trois axes de recherche sur lesquels nous espérons avoir apporté quelques contributions nouvelles :

- Analyser les fondements de la décomposition des décisions de gestion des ressources humaines dans un centre d'appels en mettant en évidence l'effet déterminant de **l'incomplétude de l'information**.
- Proposer une représentation unificatrice et un modèle générique du problème de *shift-scheduling* grâce au **paradigme de chaîne d'activités et des modèles de PLNE qui en découlent**.
- Intégrer explicitement la qualité de service dans le modèle d'optimisation du problème de *shift-scheduling* en utilisant une **approche linéaire modélisant les phénomènes stochastiques**.

Revenons sur les éléments saillants de chacune de ces trois contributions.

a) *Incomplétude de l'information et organisation des décisions*

Avant de répondre à la question « comment décider ? » qui est le propre des méthodes et modèles d'optimisation des problèmes de dimensionnement et planification, nous nous sommes d'abord intéressé à la question de « quand décider quoi ? ». C'était l'objet du second chapitre où nous avons tenté de montrer que la décomposition des décisions était rendue nécessaire par l'incomplétude de l'information. Celle-ci est liée soit au partage de la décision dans l'organisation, soit à l'incertitude des prévisions. A travers quelques exemples illustratifs, nous avons montré qu'en cas de manque ou de non fiabilité de l'information sur le futur, il était avantageux de décomposer et de prévoir des moyens de correction des décisions de gestion des ressources humaines. De ce fait, il n'existe pas d'organisation idéale des décisions valable dans l'absolu. Il appartient à chaque entreprise d'analyser son propre mode de fonctionnement et son environnement afin d'identifier les contraintes et les éléments d'incertitudes auxquels elle est soumise. L'organisation judicieuse des décisions se fera à la lumière d'une telle analyse.

b) *Paradigme de chaîne d'activités et les PLNE qui en découlent*

Dans cette thèse nous avons introduit l'idée de paradigme de chaîne d'activités. Il permet de représenter, à partir de notions simples et sans nécessiter de formalisme mathématique, la grande diversité des environnements et des contraintes de gestion des ressources humaines. Ceci confère aux modèles de PLNE qui en découlent une grande généralité pour la formulation et la résolution du problème de *shift-scheduling*. Cette vision unifiée et générique a été mise à profit à plusieurs niveaux :

- Le paradigme de chaîne d'activités a été appliqué à deux problèmes fondamentaux en gestion des ressources humaines : le dimensionnement et la planification.
- Les modèles de PLNE que nous avons construits généralisent les précédents modèles implicites de la littérature en étant totalement implicites (sur les vacations et sur les pauses).
- Nous avons intégré à nos modèles de nouveaux aspects généralement non pris en compte dans le problème de *shift-scheduling* : le multi-flux, les flux stockables, le *call-blending*, la distinction entre activités facultatives et obligatoires.

Au-delà de ces intérêts "pratiques", nous avons montré que la matrice de contraintes de temps de travail issue du paradigme de chaîne d'activités avait une structure totalement unimodulaire. Ceci nous a permis de mettre en évidence le lien entre les contraintes de couverture (dans le cas du dimensionnement) ou de qualité de service (dans le cas de la planification) et la NP-difficulté du problème de *shift-scheduling*. Nous avons également identifié quelques conditions pour lesquelles le problème était polynomial.

c) Modélisation de la qualité de service dans le problème de shift-scheduling par des approches linéaires

Dans le chapitre 4, nous nous sommes intéressés au problème de planification avec un objectif de maximisation de la qualité de service sur la journée. Pour la modélisation des contraintes de temps de travail, nous avons utilisé le modèle issu du paradigme de chaîne d'activités décrit dans le chapitre 3. En ce qui concerne l'adéquation entre la charge et la capacité, nous avons proposé deux approches linéaires pour prendre en compte la qualité de service. Une approche dite déterministe qui minimise le sous-effectif et une approche stochastique qui prend en compte les phénomènes aléatoires. A travers une application numérique, nous avons montré que cette dernière approche pouvait, sous certaines conditions, apporter une amélioration significative.

Perspectives

Un certain nombre de limitations et d'hypothèses dans nos travaux nous conduisent à envisager quelques prolongements possibles. Par exemple, nous avons considéré dans cette thèse que l'entreprise disposait des leviers d'organisation du temps de travail des conseillers. C'est la situation qui correspond à des centres d'appels internes à l'entreprise. Le recours à des prestataires étant de plus en plus courant, il serait intéressant d'étudier la possibilité de généraliser nos modèles pour intégrer ce type de problématiques.

Dans le chapitre 4, nous avons pris comme critère de qualité de service la proportion d'appels déconnectés. Nous avons supposé que les clients déconnectés ne rappellent pas. Cette supposition est généralement justifiée car les taux de rappels sont assez faibles dans la pratique. Toutefois, dans certains cas, les rappels peuvent être significatifs. Il devient alors important de les prendre en compte. Bien que ça ne figure pas ce document, nous avons étendu notre modèle pour intégrer le phénomène de rappels dans l'approche stochastique. Les contours de la méthodologie utilisée ont été présentés dans (Chauvet et Nait-Abdallah 2006; Bhulai, Koole et al. 2007; Green, Kolesar et al. 2007; Whitt 2007)

Enfin, et de manière plus fondamentale, nous pouvons envisager des développements plus "quantitatifs" en prolongement des analyses de l'organisation des décisions qui ont fait l'objet du chapitre 2. En effet, face à des problématiques d'optimisation des coûts ou de qualité de service, il est souvent pertinent de questionner le socle de (sur)contraintes que peut représenter l'organisation. L'enjeu n'est plus alors de trouver la méthode la plus optimale mais l'organisation qui permet de prendre les décisions les plus robustes face à l'incertain. Les approches d'optimisation robuste peuvent apporter des éclairages quantitatifs sur ce type de questions.

BIBLIOGRAPHIE

Aguir, S., F. Chauvet, Y. Dallery, F. Karaesmen, R. Nait-Abdallah et T. Prat (2003). Solution pour l'estimation avec intervalle de confiance du temps d'attente en vue de l'annoncer au client. Institut National de la Propriété Industrielle (INPI). France.

Altman, E., B. Gaujal et A. Hordijk (2000). "Multimodularity, convexity and optimization properties." *Mathematics of Operations Research*(25): 324-347.

Atlason, J., M. A. Epelman et S. G. Henderson (2004). "Call Center Staffing with Simulation and Cutting Plane Methods." *Annals of Operations Research* **127**(1): 333-358.

Atlason, J., M. A. Epelman et S. G. Henderson (2005). "Optimizing call center staffing using simulation and analytic center cutting plane methods." *Soumis à Management Science*(1).

Atlason, J., M. A. Epelman et S. G. Henderson (2008). "Call Center Staffing Using Simulation and Analytic Center Cutting Plane Methods." *Management Science* **54**(2): 295-309.

Aykin, T. (1996). "Optimal Shift Scheduling with Multiple Break Windows." *Management Science* **42**(4): 591-602.

Aykin, T. (2000). "A comparative evaluation of modeling approaches to the labor shift scheduling problem." *European Journal of Operational Research* **125**(2): 381-397.

Baglin, G., O. Bruel et A. Garreau (2001). *Management industriel et logistique*. Paris, Economica.

Baker, K. R. (1974). "Scheduling a Full-Time Workforce to Meet Cyclic Staffing Requirements." *Management Science* **20**(12, Application Series): 1561-1568.

Baker, K. R. (1976). "Workforce Allocation in Cyclical Scheduling Problems - A Survey." *Operational Research Quarterly* **27**(1): 155-167.

Baker, K. R. et M. J. Magazine (1977). "Workforce Scheduling with Cyclic Demands and Day-Off Constraints." *Management Science* **24**(2): 161-167.

Bartholdi, J. J. (1981). "A Guaranteed-Accuracy Round-Off Algorithm for Cyclic Scheduling and Set Covering." *Operations Research* **29**(3): 501-510.

Bechtold, S. E. et L. W. Jacobs (1990). "Implicit modeling of flexible break assignments in optimal shift scheduling." *Management Science*, **36**(11): 1339-1351.

Bechtold, S. E. et L. W. Jacobs (1996). "The equivalence of general set-covering and implicit integer programming formulations for shift scheduling." *Naval Research Logistics* **43**(2): 233-249.

Benoist, T. (2004). *Relaxations et décompositions combinatoires*. Laboratoire d'Informatique, Université d'Avignon et des Pays de Vaucluse.

Benoist, T., E. Gaudin et B. Rottembourg (2002). *Constraint Programming Contribution to Benders Decomposition: A Case Study*. Proceedings CP'02, Springer

Benoist, T., E. Gaudin et B. Rottembourg (2005). *Planification de centres d'appels téléphoniques. Gestion de production et ressources humaines Méthodes de planification dans les systèmes productifs*. V. G. Pierre Baptiste, Alain Haït, François Soumis, Presses Internationales Polytechnique (Montréal).

Bhulai, S., G. Koole et A. Pot (2007). "Simple methods for shift scheduling in multi-skill call centers." *Manufacturing & Service Operations Management* **to appear**.

Billionnet, A. (1999). "Integer programming to schedule a hierarchical workforce with variable demands." *European Journal of Operational Research* **114**(1): 105-114.

Bordoloi, S. K. et H. Matsuo (2001). "Human resource planning in knowledge-intensive operations: A model for learning with stochastic turnover." *European Journal of Operational Research* **130**(1): 169-189.

Burns, R. N. et M. W. Carter (1985). "Work force size and single shift schedules with variables demands." *Management Science* **31**(5): 599.

Canon, C. (2005). *Application des techniques de recherche opérationnelle à la planification de personnel dans un centre de contacts multi-compétent*, Université de Tours.

Cezik, M. T. et P. L'Ecuyer (2008). "Staffing Multiskill Call Centers via Linear Programming and Simulation." *Management Science* **54**(2): 310-323.

Charnes, A., W. Cooper et R. Niehaus (1978). *Management science approaches to manpower planning and organization design*, North-Holland Pub. Co.

Chauvet, F. (2005). *Une recherche opérationnelle. Habilitation à diriger des recherches. Informatique*. Paris, Université de Paris VI.

Chauvet, F., L. Chauvier, A. Renaud et R. Nait-Abdallah (2003). *An aggregate staffing model to optimize call center planning*. EURO/INFORMS Conference (ISTANBUL'2003). Istanbul (Turkey).

Chauvet, F., L. Chauvier, A. Renaud et R. Nait-Abdallah (2003). *Un modèle agrégé de planification des centres d'appels*. 5ème Congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF'2003). Avignon (France).

Chauvet, F. et R. Nait-Abdallah (2006). *Impact des renouvellements d'appels dans la planification d'un centre de clientèle*. ROADEF 2006. Lille, France.

Chauvet, F., R. Nait-Abdallah et J. Lauzier (2007). *Paradigme de chaîne d'activités pour la planification dans un centre d'appels*. FRANCORO V - ROADEF 2007. Grenoble, France.

Cleveland, B. et J. Mayben (1997). *Call center management on fast forward*, Call center press.

Coffre, A. (2005). *Délocalisation d'un centre d'appels avec voix sur IP*. 01 Réseaux. n° 149.

-
- Dantzig, G. B. (1954). "A Comment on Edie's Traffic Delays at Toll Booths." *Journal of the Operations Research Society of America* **2**(3): 339-341.
- Dowland, K. A. (1998). "Nurse scheduling with tabu search and strategic oscillation." *European Journal of Operational Research* **106**(2-3): 393-407.
- Edie, L. C. (1954). "Traffic Delays at Toll Booths." *Journal of the Operations Research Society of America* **2**(2): 107-138.
- Ernst, A. T., H. Jiang, M. Krishnamoorthy et D. Sier (2004). "Staff scheduling and rostering: A review of applications, methods and models." *European Journal of Operational Research* **153**(1): 3-27.
- Gaimon, C. et G. L. Thompson (1984). "A Distributed Parameter Cohort Personnel Planning Model that Uses Cross-Sectional Data." *Management Science* **30**(6): 750-764.
- Gans, N., G. Koole et A. Mandelbaum (2003). "Telephone call centers: Tutorial, review, and research prospects." *Manufacturing and Service Operations Management*: 5:79-141.
- Gans, N. et Y. P. Zhou (2002). "Managing learning and turnover in employee staffing." *Operations Research* **50**(6): 991-1006.
- Garey, M. R. et D. S. Johnson (1979). *Computers and intractability : a guide to the theory of NP-completeness*. San Francisco, W. H. Freeman.
- Ghouila-Houri, A. (1962). "Caractérisation des matrices totalement unimodulaires." *Comptes Rendus de l'Académie des Sciences*(254): 1192-1194.
- Ghouila-Houri, A. (1964). "Flots et tensions dans un graphe." *Annales scientifiques de l'École Normale Supérieure*(Sér. 3, 81 no. 3): p. 267-339.
- Green, L. V., P. J. Kolesar et J. Soares (2001). "Improving the SIPP Approach for Staffing Service Systems That Have Cyclic Demands." *Operations Research* **49**(4): 549-564.
- Green, L. V., P. J. Kolesar et W. Whitt (2007). "Coping with Time-Varying Demand when Setting Staffing Requirements for a Service System." *Production and Operations Management (POMS)* **16**(1): pp. 13-39.
- Grinold, R. C. et K. T. Marshall (1977). *Manpower planning models*, North Holland Publishing Company.
- Günes, E. (1999) *Workforce scheduling*. IE 672 Theory of Machine Scheduling course. Department of Industrial Engineering. Bilkent University.,
- Hajek, B. (1985). "Extremal splitting of point processes." *Mathematics of Operations Research*(10): 543-556.
- Henderson, S. G. et A. J. Mason (1998). *Rostering by Iterating Integer Programming and Simulation*. Proceedings of the 1998 Winter Simulation Conference.

Holt C., Modigliani F., Muth J. F. et S. H. (1960). Planning Production, Inventories, and Work Force, Prentice Hall, New Jersey.

ICMI (2000). Call center forecasting and scheduling : the best of call center management review, Incoming Calls Management Intitute, Call Center Press.

Ingolfsson, A., E. Cabral et X. Wu (2002). Combining Integer Programming and the Randomization Method to Schedule Employees, Department of Finance and Management Science, Faculty of Business, University of Alberta, Edmonton, Alberta, Canada.

Ingolfsson, A., M. A. Haque et A. Umnikov (2002). "Accounting for time-varying queueing effects in workforce scheduling." *European Journal of Operational Research* **139**(3): 585-597.

Jacquinet, M. (2007). Les centres de contacts, la relation clients : tout ce que vous avez toujours voulu savoir et qu'on ne vous a jamais dit, Vocalcom.

Jarrah, Bard et deSilva (1994). "Solving large scale tour scheduling problems." *Management Science* **40**(9).

Jouini, O., Y. Dallery et R. Nait-Abdallah (2004). Stochastic Models of Customer Portfolio Management in Call Centers. Proceedings of the German Operations Research Society 2004.

Jouini, O., Y. Dallery et R. Nait-Abdallah (2007). "Analysis of the Impact of Team-Based Organizations in Call Centers Management." *Management Science* **To appear**.

Kolesar, P. J., K. L. Rider, T. B. Crabill et W. E. Walker (1975). "A Queueing-Linear Programming Approach to Scheduling Police Patrol Cars." *Operations Research* **23**(6): 1045-1062.

Koole, G. (2004). Monotonicity and convexity results for multi-server queues with applications to call centers. CWI Queueing Colloquium. Amsterdam.

Koole, G. (2005). Convexity properties of queueing systems with applications to call centers INSEAD, Fontainebleau and Ecole Centrale de Paris.

Koole, G. et A. Mandelbaum (2002). "Queueing models of call centers: An introduction." *Annals of Operations Research* **113**(1-4): 41-59.

Koole, G. et E. van der Sluis (2003). "Optimal Shift Scheduling with a Global Service Level Constraint." *IIE Transactions* **35**(11): 1049 - 1055.

Mabert (1979). "A Case Study of Encoder Shift Scheduling under Uncertainty." *Management Science* **25**(7).

Mabert, V. A. et C. A. Watts. (1982). "A simulation analysis of tour-shift scheduling construction procedures." *Management Science* **28**(5).

Moondra, S. L. (1976). "An L.P. model for work force scheduling for banks." *Journal of Bank Research* **7**(4): 299-301.

Morris, J. G. et M. J. Showalter (1983). "Simple approaches to shift, days-off and tour scheduling problems." *Management Science* **29**.

Nagarajan, R. et D. Towsley (1992). A note on the convexity of the probability of a full buffer in the M/M/1/K queue, CMPSCI Technical Report TR 92-85.

Partouche, A. (1998). Planification d'horaires de travail. Méthodologie, modélisation et résolution à l'aide de la programmation linéaire en nombres entiers et de la programmation par contraintes. U.F.R Sciences des organisations. Paris, Université Paris-Dauphine.

Rekik, Soumis et Cordeau (2003) Implicit Shift Scheduling with Multiple Breaks and Pre- and Post-Break Restrictions. Cahiers du Gerad.,

Rekik, M., J. F. Cordeau et F. Soumis (2004). "Using Benders decomposition to implicitly model tour scheduling." *Annals of Operations Research* **128**(1-4): 111-133.

Reynolds, P. (2003). Call center staffing : the complete, practical guide to workforce management, The Call Center School.

Reynolds, P. (2003). Call center staffing. The complete, practical guide to workforce management, The Call Center School.

Rottembourg, B. (2003). Dimensionnement et planification des centres d'appels : problématiques et solutions d'optimisation. 4e conférence francophone de Modélisation et de Simulation MOSIM'03. Toulouse.

Segal, M. (1974). "Operator-Scheduling Problem - Network-Flow Approach." *Operations Research* **22**(4): 808-823.

Thompson, G. M. (1993). "Accounting for the multi-period impact of service when determining employee requirements for labor scheduling." *Journal of Operations Management* **11**(3): 269-287.

Thompson, G. M. (1995). "Improved Implicit Optimal Modeling of the Labor Shift Scheduling Problem." *Management Science* **41**(4): 595-607.

Thompson, G. M. (1997). "Labor staffing and scheduling models for controlling service levels." *Naval Research Logistics* **44**(8): 719-740.

Thompson, G. M. et J. C. Goodale (2006). "Variable employee productivity in workforce scheduling." *European Journal of Operational Research* **170**(2): 376-390.

Tien, J. M. et A. Kamiyama (1982). "On Manpower Scheduling Algorithms." *Siam Review* **24**(3): 275-287.

Wallace, R. B. et W. Whitt (2005). "A staffing algorithm for call centers with skill-based routing." *Manufacturing & Service Operations Management* **7**(4): 276-294.

Warner, D. M. (1976). "Scheduling Nursing Personnel According to Nursing Preference - Mathematical-Programming Approach." *Operations Research* **24**(5): 842-856.

Whitt, W. (2007). "What You Should Know About Queueing Models to Set Staffing Requirements in Service Systems." *Naval Research Logistics* **54**(5): 476-484.

Résumé

Cette thèse aborde la gestion des ressources humaines dans un centre d'appels. Plus spécifiquement, nous nous intéressons aux problèmes de dimensionnement et de planification. L'objectif sous-jacent est d'assurer la meilleure qualité de service au client (par exemple minimiser le délai d'attente) avec un coût salarial minimum pour l'entreprise. Ces problématiques sont généralement modélisées dans la littérature par le problème de construction de vacation (*shift-scheduling problem*). Pour appréhender ce problème, nous introduisons le paradigme de chaîne d'activités. Ce paradigme nous permet de représenter la grande diversité des environnements et des contraintes de gestion des ressources humaines dans un centre d'appels. Nous traduisons ensuite ce paradigme en programme linéaire en nombres entiers pour résoudre les problèmes de dimensionnement et de planification. Nous proposons enfin une méthode pour intégrer au programme linéaire en nombres entiers un objectif de qualité de service non linéaire.

Mots clés : centres d'appels, planification, dimensionnement, problème de construction de vacances, programmation linéaire en nombres entiers, modèles implicites.

Abstract

We address in this thesis dimensioning and planning issues in a call center. The purpose is to guarantee a high quality of service at a minimum cost for the company. In the literature, these issues are generally modeled using the shift-scheduling problem. In this thesis we introduce what we call the activity series paradigm. This paradigm allows dealing with the great diversity of call center constraints and environments. It was used to model and solve the planning and dimensioning problems with an integer linear program. We also develop a method to optimize a non-linear quality of service in the shift-scheduling problem.

Keywords: call centers, planning, dimensioning, shift-scheduling problem, integer linear programming, implicit models.