

N° d'ordre: 3242

# THÈSE

présentée

devant l'université de Rennes 1

pour obtenir

le grade de : DOCTEUR DE L'UNIVERSITÉ DE RENNES 1  
Mention : INFORMATIQUE

par

Yezekael HAYEL

Équipe d'accueil : ARMOR  
École doctorale : MATISSE  
Composante universitaire : IRISA

Titre de la thèse :

Tarifcation des réseaux à différenciation de services

Soutenue le 30 Novembre 2005 devant la commission d'examen

Mme	Nihal	PEKERGIN	Rapporteurs
M.	Eitan	ALTMAN	
Mme	Luce	BROTCORNE	Examineurs
M.	Raymond	MARIE	
M.	Bruno	TUFFIN	Encadrant
M.	Gerardo	RUBINO	Directeur
M.	Denis	PHAN	Invité



à ma femme Laurence,



## Remerciements

Je tiens tout d'abord à exprimer toute ma gratitude envers mon encadrant de thèse Bruno Tuffin pour ses nombreuses aides. Je le remercie tout particulièrement pour ses conseils, sa disponibilité et son soutien qui m'ont aidé à avancer dans ma thèse. Je remercie également Gerardo Rubino de m'avoir accepté au sein de son projet et de m'avoir fait confiance depuis le début pour diriger mes travaux.

Je remercie Nihal Pekergin et Eitan Altman qui ont accepté de rapporter ce manuscrit et de faire le déplacement jusqu'en Bretagne pour être présents lors de la soutenance.

Un grand merci à Raymond Marie pour avoir présidé mon jury de soutenance et de sa confiance qu'il m'a adressée pour les travaux dirigés de Modélisation et Files d'attente.

Merci à Luce Brotcorne et Denis Phan qui ont accepté de participer au jury de soutenance et pour l'intérêt qu'ils ont manifesté pour mes travaux.

Je remercie Laura Wynter et Parijat Dube d'IBM T.J. Watson Research Center pour leur accueil chaleureux durant mon stage et pour ces trois mois passés au sein de leur équipe.

Je voudrais également remercier l'ensemble de l'équipe ARMOR pour leur bonne humeur et leur convivialité. Un grand merci à Martin et Francine pour avoir partagé au quotidien l'espace de travail, ce qui nous a permis de nous apprécier bien au delà du contexte du travail.

Merci à tous les copains pour les soirées de discussion sans fin, je pense à Antoine, Julien, Fred, Aurel, Morgan et pardon à ceux que j'oublie. Je remercie aussi toute ma famille, mes parents et mes soeurs de leur soutien moral.

Finalement, je remercie tout particulièrement les deux femmes de ma vie Laurence et Lazuli, pour leur soutien inconditionnel et leur amour qui est le meilleur des encouragements.



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>9</b>
1.1	La tarification des réseaux de communication . . . . .	10
1.2	Tarification et gestion de la qualité de service . . . . .	11
1.3	Outils d'analyse des modèles de tarification de l'Internet . . . . .	12
1.4	État de l'art . . . . .	13
1.5	Contributions . . . . .	14
<b>2</b>	<b>Modèle de tarification du <i>Cumulus</i> pour l'architecture <i>DiffServ</i></b>	<b>19</b>
2.1	Présentation du modèle du <i>Cumulus</i> . . . . .	20
2.2	Principes de fonctionnement . . . . .	21
2.3	Définitions . . . . .	23
2.3.1	Le revenu du fournisseur . . . . .	23
2.3.2	Satisfaction de l'utilisateur . . . . .	24
2.4	Problème d'incitation . . . . .	25
2.5	Le modèle du <i>Cumulus</i> avec pénalisation . . . . .	27
2.5.1	Définitions . . . . .	27
2.5.2	Problème d'incitation . . . . .	28
2.6	Optimisation du revenu du fournisseur . . . . .	30
2.6.1	Minimisation du coût du fournisseur . . . . .	30
2.6.2	Optimisation avec configurations particulières . . . . .	31
2.7	Applications numériques . . . . .	35
2.7.1	Compatibilité d'incitation . . . . .	36
2.7.2	Optimisation dans le cas général . . . . .	36
2.7.3	Optimisation avec configurations particulières . . . . .	39
2.8	Conclusion . . . . .	41
<b>3</b>	<b>Mécanismes d'ordonnancement et tarification</b>	<b>45</b>
3.1	Modélisation . . . . .	47
3.1.1	Modèle de trafic . . . . .	47
3.1.2	Modèle des utilisateurs . . . . .	48
3.2	Politique de service avec priorités strictes . . . . .	49
3.3	Politique de service à processeur partagé généralisé (GPS) . . . . .	49
3.3.1	Optimisation du revenu du fournisseur . . . . .	51

3.3.2	Comparaison des revenus optimaux avec GPS et avec priorités strictes . . . . .	60
3.4	Politique de service à processeur partagé de façon discriminante (DPS) . . . . .	63
3.4.1	Jeu non-coopératif entre classes de service . . . . .	64
3.4.2	Étude de l'équilibre du système . . . . .	65
3.4.3	Optimisation du revenu du fournisseur . . . . .	69
3.4.4	Comparaison des revenus optimaux avec DPS et priorités strictes . . . . .	71
3.4.5	Application aux trafics TCP . . . . .	72
3.5	Conclusion . . . . .	73
<b>4</b>	<b>Mécanisme de partage du coût de congestion et priorités strictes</b>	<b>75</b>
4.1	Modélisation . . . . .	76
4.2	Temps de service homogènes . . . . .	78
4.2.1	Tarifcation optimale . . . . .	78
4.2.2	Propriété de compatibilité d'incitation . . . . .	81
4.3	Temps de service non-homogènes . . . . .	84
4.4	Tarifcation dynamique mono-classe . . . . .	87
4.5	Extension à un réseau de Jackson ouvert . . . . .	90
4.6	Conclusion . . . . .	95
<b>5</b>	<b>Tarifcation avec mesures actives et passives des performances</b>	<b>97</b>
5.1	Modélisation . . . . .	98
5.2	Tarifcation Optimale . . . . .	99
5.3	Mesures actives et passives des performances . . . . .	100
5.3.1	Estimation du débit . . . . .	101
5.3.2	Estimation du délai moyen . . . . .	102
5.4	Procédure d'estimation des performances . . . . .	107
5.5	Algorithmes d'ajustement des paramètres de mesures . . . . .	108
5.6	Applications numériques . . . . .	109
5.6.1	Valuation constante . . . . .	109
5.6.2	Valuation variable . . . . .	111
5.7	Conclusion . . . . .	112
<b>6</b>	<b>Tarifcation dans un réseau sans fil CDMA</b>	<b>115</b>
6.1	Modèle de transmission CDMA . . . . .	116
6.1.1	Mesures de performance . . . . .	117
6.1.2	Contrainte de capacité . . . . .	118
6.2	Modélisation des utilisateurs . . . . .	118
6.3	Jeu non-coopératif entre classes de service . . . . .	119
6.4	Tarifcation optimale avec une classe de service . . . . .	120
6.4.1	Demande infinie . . . . .	120
6.4.2	Demande aléatoire . . . . .	126
6.5	Tarifcation optimale multi-services . . . . .	129
6.5.1	Demande infinie . . . . .	130



---

6.5.2	Demande aléatoire . . . . .	137
6.6	Conclusion . . . . .	141
<b>7</b>	<b>Allocation optimale avec choix aléatoire de la classe de service</b>	<b>143</b>
7.1	Le <i>Yield Management</i> et la tarification des réseaux . . . . .	143
7.2	Modèle de choix discret aléatoire du Logit . . . . .	144
7.3	Optimisation du revenu moyen du fournisseur . . . . .	146
7.4	Application à des requêtes sur un serveur web . . . . .	151
7.5	Conclusion . . . . .	152
<b>8</b>	<b>Conclusion générale</b>	<b>155</b>
8.1	Contributions . . . . .	155
8.2	Perspectives . . . . .	157



# Chapitre 1

## Introduction

L'Internet est de plus en plus présent dans nos foyers, on comptait une connexion pour 2 foyers en France en 2003 avec une croissance prévisionnelle de 30% pour les 5 années à venir<sup>1</sup>. D'un point de vue mondial, le trafic sur le cœur de l'Internet a doublé tous les six mois entre 1997 et 2001 [1]. L'engouement des français pour les connexions haut-débit, ADSL ou câble, conforte cette tendance. De plus, les perspectives de nouvelles applications telles la téléphonie sur Internet ou la vidéo à la demande vont accentuer l'importance du réseau Internet dans les foyers comme un bien de consommation accessible à tout le monde.

Le développement rapide de l'Internet a été facilité par la simplicité du modèle de tarification proposé par les fournisseurs d'accès. En effet, cette tarification est forfaitaire et indépendante de la consommation. Elle est nommée dans ce cas tarification «plate» (*flat rate pricing* [79]). Elle n'est pas équitable pour les utilisateurs car les gros consommateurs qui représentent 20% des utilisateurs exploitent 80% des ressources sans payer plus cher que les petits consommateurs. D'autre part, l'accroissement rapide du nombre d'utilisateurs et de leur trafic entraîne des situations de congestion de plus en plus fréquentes. Dès lors, l'un des problèmes majeurs de l'Internet est le contrôle de la congestion. Une solution à ce problème serait l'extension de l'utilisation de fibres optiques, qui permettent des débits très importants [9]. Néanmoins dans les réseaux d'accès, l'installation de ces fibres optiques est très coûteuse. Ce problème est connu sous le nom de problème du *last mile* [17].

L'un des principaux mécanismes pour contrôler la congestion dans l'Internet actuel se fait au niveau de la couche transport via le protocole de transfert TCP [4] (*Transmission Control Protocol*). Celui-ci contrôle les débits en fonction de la congestion, mais sans s'occuper de la *demande* des utilisateurs. Dans des situations de congestion du réseau, contrôler la *demande* de l'utilisateur permettrait d'assurer une meilleure qualité de service. De plus, les utilisateurs n'évaluent pas de la même façon le service perçu. Certains seraient prêts à payer plus cher afin d'obtenir une meilleure qualité de

---

<sup>1</sup>(source : <http://www.internet.gouv.fr>) Ce site du gouvernement propose une synthèse des principales études quantitatives proposées par les instituts de sondage et organismes de la société de l'information.

service.

Les réseaux sans fil sont également touchés par ce problème de congestion. En effet, leur capacité de transmission est limitée par le spectre des fréquences radio. Par exemple, les réseaux sans fil de troisième génération utilisant la technologie CDMA (*Code Division Multiple Access*) qui proposent de nouvelles applications temps-réel comme la vidéo-conférence ou la télévision (qui requièrent une grande quantité de ressources), ne pourront pas offrir un débit maximal à tous les utilisateurs simultanément.

Nous proposons dans cette thèse d'utiliser des mécanismes de tarification qui se basent sur la relation inhérente entre l'offre et la demande pour contrôler la congestion dans les réseaux de communication. On peut noter que des mécanismes de tarification sont également utilisés pour contrôler le trafic dans d'autres types de réseaux comme par exemple dans les réseaux de transports [8, 25, 98] ou les réseaux aériens [36].

## 1.1 La tarification des réseaux de communication

La construction d'un modèle de tarification doit prendre en compte les caractéristiques et les envies des acteurs économiques de l'Internet qui sont d'un côté les fournisseurs d'accès ou de services et de l'autre les utilisateurs. Afin de connaître la réaction des utilisateurs face à de nouveaux mécanismes de tarification, des études socio-économiques ont été réalisées dans le projet INDEX (*Internet Demand Experiment* [41]). Cette étude montre que les utilisateurs portent un intérêt particulier à connaître au préalable une idée de leur coût financier. Des expérimentations réelles de tarification à l'usage ont été réalisées dans certains pays, notamment pour le réseau universitaire en Nouvelle-Zélande [49].

Quant au fournisseur de service, le nouveau modèle de tarification doit lui permettre de recouvrir ses coûts. En effet, la mise en service d'une architecture de réseau différente et tous les besoins pour la mise en place d'un nouveau système de tarification (collection des données, mesures de trafic, nouvelles applications,...) induisent des investissements financiers importants de la part du fournisseur. Certains chercheurs se penchent justement sur ce problème de définition des coûts pour un fournisseur de service [95].

On peut donc résumer les deux principaux arguments en faveur d'un nouveau modèle de tarification à l'usage :

- pouvoir contrôler la demande et donc la congestion en utilisant la relation inhérente entre Demande/Prix (voir Figure 1.1(a)),
- pouvoir assurer un revenu pour le fournisseur (voir figure 1.1(b)).

Dans cette thèse, nous étudions des mécanismes de tarification à l'usage statique où les prix sont fixés. Cette tarification est moins efficace qu'une tarification dynamique mais il a été montré dans [41, 105] que les utilisateurs préfèrent avoir une idée a priori de leur facture, ce qui est le cas avec une tarification statique et non avec une tarification dynamique.

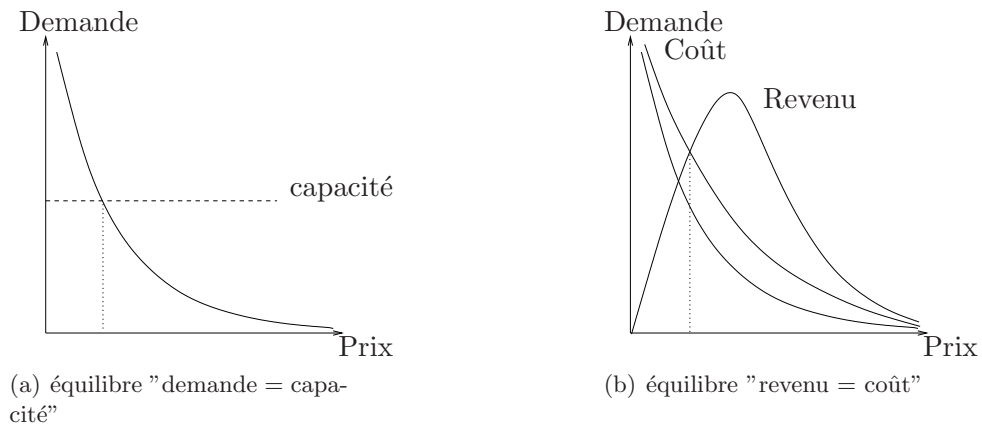


FIG. 1.1 – Les objectifs d’une tarification à l’usage des réseaux : contrôle de congestion et retour sur investissement du fournisseur.

## 1.2 Tarification et gestion de la qualité de service

L’Internet actuel est basé sur l’architecture de réseau IP (pour *Internet Protocol*). Le protocole IP fournit un simple service où tous les datagrammes sont transmis sans aucune garantie de délai ou d’assurance de la livraison des datagrammes. L’émergence de nouvelles applications temps-réel telles la téléphonie sur Internet (*VoIP* pour *Voice Over IP*) ou vidéo à la demande (*VoD* pour *Video on Demand*) avec des demandes importantes en terme de performance nécessite une adaptation du réseau actuel. L’*Internet Engineering Task Force* (IETF) est une organisation internationale regroupant chercheurs, ingénieurs et opérateurs de télécommunication. Elle s’intéresse à l’évolution de l’architecture de l’Internet. Deux groupes de travail ont proposé une architecture de réseau permettant de gérer la qualité de service sur Internet.

- La première architecture utilise une réservation de ressource point-à-point pour chaque flux de transmission. Cette architecture de réseaux, appelée *IntServ* [23], ne résiste pas au facteur d’échelle et peut donc difficilement être mise en place sur l’Internet.
- La seconde architecture basée sur une différenciation de service, permet de promouvoir de la qualité de service sans réservation de ressources. Cette architecture, nommée *DiffServ* [114], considère une agrégation de trafic en classes de service traitées différemment aux routeurs internes du domaine *DiffServ*. Le comportement EF (pour *Expedited Forwarding*) est un exemple de classe de service standardisée sur ce type d’architecture de réseau qui obtient une meilleure qualité de service (moins de délai et de perte) que les autres classes via des mécanismes de priorités au sein des routeurs du réseau.

Les utilisateurs d’Internet ayant un comportement égoïste où chacun cherche à optimiser sa qualité de service, si le fournisseur ne différencie pas les classes de service par une tarification ou toute autre technique de contrôle adaptée, chacun aura intérêt à choisir la classe de service fournissant la meilleure qualité [73]. Ainsi, le problème de

congestion totale du réseau se réduirait à la congestion d'une classe de service et aucune différenciation ne sera réellement appliquée. La gestion de la qualité de service dans les réseaux avec la mise en place de mécanismes de différenciation de services est donc fortement liée à son modèle de tarification [96].

La gestion de la qualité de service dans les réseaux sans fil est également très importante. En effet, pour ce type de réseau, la ressource est limitée et donc les phénomènes de congestion importants. Des mécanismes de différenciation de services ont été proposés pour ces types de réseaux, par exemple pour les réseaux ad-hoc, la norme 802.11e propose de la différenciation de services à la *DiffServ* [115]. Là encore, l'introduction de la tarification est un moyen simple pour faire du contrôle sur ce type de réseau.

### 1.3 Outils d'analyse des modèles de tarification de l'Internet

Dans les réseaux de communication, la satisfaction des utilisateurs dépend de la qualité du service perçue. Celle-ci s'exprime en fonction de différentes métriques comme le délai moyen, la gigue (variation de délai), la probabilité de perte et le débit. Nous considérons dans cette thèse la métrique du délai moyen, comme dans de nombreux modèles de tarification de la littérature. En effet, la plupart des applications de l'Internet (transfert de données, WEB, applications multimédia,...) sont sensibles à ce paramètre de qualité de service. Le délai de bout-en-bout est dû au traitement et à l'attente de service des paquets dans les routeurs. Cette attente de service dépend de l'intensité du trafic qui traverse le routeur. Ainsi, un utilisateur qui envoie ses paquets à travers le réseau induit un coût appelé en économie externalité, reflétant l'augmentation du délai pour les paquets des autres utilisateurs.

Un mécanisme de tarification possède différents critères économiques d'efficacité. Les deux principaux que l'on retrouve dans la littérature sont : maximisation du surplus collectif (bien-être du système) ou maximisation du revenu du fournisseur. Afin de montrer ces propriétés, nous devons appréhender le comportement des utilisateurs. Ceux-ci sont supposés égoïstes, c'est-à-dire que chacun agit de façon à optimiser son utilité sans prendre en considération l'externalité que cela implique sur les autres. Le cadre théorique pour l'étude de tels systèmes est la théorie des jeux et plus précisément les jeux non-coopératifs. Suivant le mécanisme de tarification, nous montrons l'existence d'équilibres de Nash [45]. Ce type d'équilibre où aucun joueur ne peut unilatéralement augmenter son utilité est souvent recherché pour les jeux non-coopératifs. Dans cette thèse, l'analyse d'un jeu non-coopératif se fait en observant sa dynamique à plusieurs niveaux. Nous utilisons dans les chapitres 3 et 6, un modèle de jeu à deux niveaux liés à l'équilibre de Stackleberg [12].

- Les utilisateurs prennent leur décision afin d'optimiser leur utilité en fonction des paramètres du mécanisme de tarification. Les paramètres des utilisateurs peuvent être la classe de service, le débit, etc.
- Le système détermine les paramètres du modèle comme les prix, les allocations, etc. qui optimisent le surplus collectif ou le revenu.

Une notion importante en théorie des jeux est de pouvoir évaluer quantitativement la satisfaction des utilisateurs et cette notion est primordiale pour l'analyse des modèles de tarification des réseaux. Les préférences (ou niveau de satisfaction) des utilisateurs s'expriment via une fonction d'*utilité*. Cette fonction dépend généralement de la qualité de service et du type d'application dans un contexte de trafics hétérogènes. Nous avons considéré dans cette thèse des utilités données sous forme de fonctions explicites du délai moyen, lequel pouvant être obtenu sous une forme explicite suivant le modèle de réseau proposé. Ces fonctions sont supposées différentes en fonction du type d'application ou de trafic. Dans la littérature, les fonctions d'utilité sont choisies d'une part en accord avec des études sur le comportement des utilisateurs [46] et d'autre part pour satisfaire des propriétés mathématiques intéressantes comme continuité, dérivabilité et concavité qui permettent d'assurer l'existence et l'unicité de points optimaux [67].

## 1.4 État de l'art

Depuis une dizaine d'années, le domaine de la tarification des réseaux de communication a inspiré de nombreux travaux. Une classification exhaustive des modèles est difficile en raison de la grande diversité des mécanismes économiques (enchères, tarification à l'usage ou encore au débit) et d'implémentations (tarification par paquet/flux, trafic montant/descendant ou encore tarification à chaque nœud/routeur frontière/réseaux d'accès,...). On s'intéresse principalement dans cette thèse aux modèles sans réservation de ressource adaptés à une architecture de réseaux à différenciation de services laquelle est envisagée pour le futur Internet afin de promouvoir une meilleure qualité de service que le réseau actuel.

Les mécanismes simples de tarification des réseaux définissent les prix afin de contrôler le débit de transmission des utilisateurs. Concrètement, l'utilisateur définit le prix qu'il est prêt à payer par unité de ressource (débit). Ensuite, le système détermine de façon optimale le débit associé à chaque utilisateur. Ces modèles ont été utilisés particulièrement pour les trafics élastiques dans [58]. Ce type de tarification a été également proposé pour tarifier la classe de service ABR (pour *Available Bit Rate*) dans un contexte de réseau ATM (pour *Asynchronous Transfer Mode*) [31]. De nombreuses extensions ont été proposées dans [59, 60] ou encore [106]. D'autres modèles ont été construits suivant le mécanisme inverse. Les utilisateurs définissent le taux de transmission qu'ils désirent et le réseau détermine de façon optimale le prix [71, 72].

Un modèle plus simple est celui dit du métro parisien (PMP pour *Paris Metro Pricing*) décrit dans [86]. Ce mécanisme propose de segmenter la demande totale en plusieurs classes. Les classes sont strictement identiques en termes de service, seuls les prix diffèrent. Ainsi, les classes les plus chères seront moins congestionnées que les autres. Mais, ce modèle simple possède plusieurs inconvénients. Dans [47], Gibbens et al. montrent que ce modèle n'est pas stable dans un environnement concurrentiel entre fournisseurs d'accès. Dans [27], Cao et Shen comparent ce modèle avec un modèle à priorités strictes et montrent que ce dernier engendre un meilleur surplus collectif que PMP. Finalement, dans [97], Tuffin et Ros ont montré que PMP engendre un revenu

moins important pour le fournisseur que s'il n'y avait qu'une seule classe de service. Malgré tous ces inconvénients, ce modèle reste très attractif car simple à mettre en œuvre sur un réseau à différenciation de service.

Ce dernier modèle ne prend pas en compte le mécanisme d'ordonnement entre les classes de service. Un algorithme d'ordonnement largement étudié est le mécanisme à priorités strictes entre les classes de service. Des modèles proposent d'étudier une tarification suivant les niveaux de priorité des classes de service. Nous avons noté dans la littérature plusieurs façons de définir le prix associé à la classe de priorité.

- Nous avons des modèles avec des priorités dont le prix est fixé à l'avance. Par exemple, dans [29] le prix de chaque classe de priorités est fixé à l'avance et l'ordre des prix suit l'ordre des priorités. Dans [74, 76], Marbach propose un modèle mathématique et montre l'existence d'un équilibre entre les demandes des utilisateurs.
- Il existe d'autres modèles de tarification pour des systèmes multiclassés mais dans lesquels le prix de chaque classe de priorité n'est pas fixé et dépend des performances du trafic comme le délai moyen [81]. Ce dernier mécanisme, proposé par Mendelson et Whang, est basé sur une tarification au coût marginal, où chaque utilisateur est facturé du coût perçu par les autres du fait de sa présence dans le système.

L'ensemble des modèles précédents sont des mécanismes dits de tarification statique car le prix est défini et fixé pour une période temps assez grande. Un compromis dans la fréquence des mises à jour des prix doit être déterminé. Une granularité (fréquence de mise-à-jour des prix) trop petite n'est pas appréciable pour les utilisateurs [32] car les prix varient souvent et l'utilisateur est dans l'impossibilité de prévoir sa facture. Une granularité trop grande est moins efficace pour le revenu du fournisseur. Une étude de ce problème est réalisée dans [89]. Les auteurs montrent finalement qu'une tarification statique avec des variations à l'échelle de la journée (*Time-of-day*) est un bon compromis.

Dans cette thèse nous étudions des modèles de tarification à l'usage afin de faire du contrôle. Dans [102], les auteurs proposent un compromis entre complexité et efficacité par rapport au modèle de la tarification plate et de la tarification à l'usage. Le prix dépend de la future congestion le long du chemin utilisé par les paquets. Toutes les informations sur les consommations et utilisations de ressources nécessaires à la détermination du prix, doivent être connues aux frontières du domaine *DiffServ*. Le modèle du Cumulus [94], étudié dans le chapitre 2, fait parti de cette famille de modèles.

La liste précédente des différents modèles est loin d'être exhaustive mais donne un aperçu de la richesse du domaine. De nombreuses études bibliographiques ont été proposées dans la littérature sur ce sujet, on peut citer par exemple [34, 37, 109].

## 1.5 Contributions

Nous étudions tout d'abord le mécanisme du Cumulus (*Cumulus Pricing Scheme*) proposé dans [94]. Ce modèle de tarification à l'usage peut être mis en place aisément



dans l'architecture de réseaux *DiffServ* par le biais de contrats (SLA pour *Service Level Agreement*) entre utilisateurs et fournisseurs [107]. Il possède de nombreuses propriétés économiques intéressantes et son principal intérêt est de prévoir la consommation moyenne des utilisateurs sur une période de temps donnée. En effet, ce modèle utilise un contrat entre utilisateur et fournisseur d'accès portant sur la quantité de ressources consommée sur une période de temps. Un mécanisme de pénalisation est mis en place afin d'inciter les utilisateurs à respecter la quantité spécifiée dans leur contrat. Le fournisseur connaît alors la consommation moyenne de chaque utilisateur sur une période donnée. Nous montrons qu'il est nécessaire de modifier le modèle initial du Cumulus afin d'obtenir cette propriété de *compatibilité d'incitation*. Ensuite, nous déterminons les valeurs de paramètres afin que le revenu moyen du fournisseur soit optimisé sous l'hypothèse de *compatibilité d'incitation*. Le modèle du Cumulus est intéressant pour les réseaux à différenciation de service comme *DiffServ* car il est simple à mettre en œuvre sur cette architecture de réseau, et il ne dépend pas des mécanismes de gestion du trafic.

Dans le chapitre 3, on s'intéresse justement à l'impact économique des mécanismes de différenciation de service mis en place dans les routeurs d'un domaine *DiffServ*. Nous étudions particulièrement différentes politiques d'ordonnancement proposées dans les standards. Ces mécanismes ont un impact important sur la qualité de service et il est donc intéressant de les étudier et de les comparer dans un contexte économique. Nous considérons les deux principaux mécanismes d'ordonnancement mis en œuvre sur ce type d'architecture de réseaux : les priorités strictes et la famille des modèles *Weighted Fair Queueing* [35]. Ce dernier est une émulation par paquet du modèle fluide du processeur partagé généralisé (GPS pour *Generalized Processor Sharing* décrit dans [88]). Nous modélisons le réseau ou un lien congestionné par une file M/M/1 et observons que le fournisseur optimise son revenu lorsqu'il propose un mécanisme à différenciation de services. Ensuite, nous comparons le revenu maximum du fournisseur obtenu avec les politiques d'ordonnancement à priorités strictes et GPS. Dans un contexte de trafic important, le mécanisme GPS peut être approché par un système de multiple files d'attente indépendantes en parallèle pour lequel le délai moyen est explicitement connu. Nous montrons alors que le mécanisme d'ordonnancement à priorités strictes donne un meilleur revenu que le modèle de GPS approché. On s'intéresse ensuite à l'algorithme d'ordonnancement à processeur partagé de façon discriminatoire (DPS pour *discriminatory processor sharing*). Ce dernier est largement étudié dans la littérature ([11, 110]) car il modélise le comportement de connexions TCP en compétition à un noeud du réseau. Nous étudions l'impact des paramètres de ce mécanisme sur le revenu du fournisseur et observons également que le revenu maximum est obtenu en considérant des priorités strictes. Ainsi, dans un contexte de réseau à différenciation de service, le fournisseur améliore son revenu en imposant une politique de service à priorités strictes par rapport à GPS. De plus, si l'on considère des connexions TCP en concurrence, le mécanisme à priorités strictes donne également un meilleur revenu pour le fournisseur.

Il existe de nombreux modèles de tarification dans la littérature qui se basent sur le mécanisme d'ordonnancement à priorités strictes. Beaucoup de modèles proposent une tarification au coût marginal qui ne partage pas le coût total de congestion du

système. Nous proposons alors dans le chapitre 4, un mécanisme de tarification qui partage ce coût total de congestion pour un système à différenciation de service avec priorités strictes. Le mécanisme proposé est optimal au sens du surplus collectif. Nous proposons ensuite une extension dynamique de ce nouveau modèle où la demande varie au cours du temps puis, une extension à un réseau ouvert de files d'attente.

Ce mécanisme suppose que le débit optimal donné par le modèle correspond à la consommation réelle des utilisateurs, ce qui n'est pas forcément vérifié suivant la pertinence du modèle. Nous proposons donc dans le chapitre 5, que le système réalise des mesures sur le réseau afin de connaître les débits réels des utilisateurs. La métrologie des réseaux consiste à analyser et à caractériser le trafic par des mesures effectuées directement sur le système. Nous étudions un mécanisme de tarification à l'usage dans ce contexte de mesures de performances du réseau. Nous déterminons les paramètres du fournisseur et de mesure qui optimisent le surplus collectif du système.

Les réseaux sans fil sont également fortement sujets au problème de congestion. Dans le chapitre 6, nous nous intéressons à un modèle de tarification à l'usage afin de contrôler le trafic et faire de la différenciation de service dans les réseaux sans fil et en particulier les réseaux de troisième génération de type UMTS (*Universal Mobile Telecommunications System*) basés sur la technologie CDMA [113]. Ce type de réseau est particulièrement sensible au problème de congestion car la capacité du spectre des fréquences radio est limitée dans les normes [55]. Nous étudions l'impact sur le revenu du fournisseur d'une tarification à l'usage pour les transmissions sur la voie montante (des mobiles vers la station de base). La technologie CDMA permet de desservir plusieurs utilisateurs sur la même bande de fréquence, sur la même période de temps. Nous nous limitons à l'étude d'une seule cellule et nous considérons que la qualité de service des utilisateurs dépend uniquement des interférences intra-cellulaires. Nous considérons plusieurs types d'utilisateurs (représentant plusieurs types de services ou applications) et déterminons les paramètres du modèle qui optimisent le revenu de la station de base. L'originalité de ce modèle de tarification est de considérer une tarification statique préférée des utilisateurs face aux mécanismes dynamiques proposés dans la littérature, et qui prend également en compte un mécanisme de contrôle de puissance optimale.

Finalement, le dernier chapitre présente une étude d'un modèle de tarification reliant l'optimisation du revenu à l'allocation de la ressource. Ces problèmes d'optimisation sont connus en marketing sous le nom de tarification en temps-réel ou *yield management*. Cette technique permet de calculer les meilleurs prix pour optimiser le profit généré par la vente d'un produit ou service, sur la base d'une modélisation et d'une prévision du comportement de la demande par micro-segment de marché. La segmentation consiste à découper le marché en sous-ensembles homogènes du point de vue des besoins et est susceptible de constituer des marchés potentiels distincts. Le *Yield Management* est utilisé depuis les années 80 dans les réseaux aériens [15] et permet par exemple, à la compagnie American Airlines de générer 1,4 milliards de dollars de revenu supplémentaire entre 1989 et 1991 grâce à son utilisation systématique pour déterminer le nombre de sièges alloué à chaque classe de prix (segment du marché)<sup>2</sup>.

---

<sup>2</sup>source : [http://www.optims.com/hight\\_profits.html](http://www.optims.com/hight_profits.html)

Nous considérons différentes classes de prix (segmentation du marché) et nous déterminons la quantité de ressources à allouer pour chacune des classes afin d'optimiser le revenu moyen du système.

Nous concluons la thèse dans le chapitre 8 et donnons quelques perspectives de recherche pour nos travaux futurs.



## Chapitre 2

# Modèle de tarification du *Cumulus* pour l'architecture *DiffServ*

Le modèle du *Cumulus* a été introduit par Reichl et al. dans [94]. Les auteurs ont montré qu'il répondait à de nombreux critères d'un « bon » modèle de tarification, principalement en raison de sa simplicité de mise en œuvre sur un réseau à différenciation de service de type *DiffServ*. Ce modèle est également apprécié des utilisateurs car leur facture est prévisible. De plus, ce mécanisme de tarification possède de nombreuses propriétés d'efficacité [93].

Le modèle du *Cumulus* s'intègre dans une architecture de réseaux à différenciation de service telle que *DiffServ* car il opère aux frontières du domaine (routeurs où est mesurée la consommation des utilisateurs). Dans *DiffServ*, des contrats, appelés *SLA* (pour *Service Level Agreement*), sont passés entre fournisseur et utilisateurs portant sur les caractéristiques propres du trafic de l'utilisateur. Les paramètres du modèle du *Cumulus* peuvent également faire partie du même *SLA*. Ainsi, le contrat détermine quelles sont les classes de service choisies par l'utilisateur et la tarification associée. Il faut noter qu'il existe aussi des *SLA* entre domaines *DiffServ* (entre fournisseurs), qui sont négociés sur la base des estimations de trafic inter-domaines et le modèle du *Cumulus* peut y être appliqué [107].

Ce chapitre porte sur l'analyse des performances de ce modèle d'un point de vue économique (et a été publié dans [53]). Pour cela, nous modélisons le *Cumulus* pour chacun des acteurs sur un domaine du réseau : les utilisateurs et le fournisseur de service. Le contrat entre les utilisateurs et leur fournisseur porte sur la quantité de ressource allouée sur une période de temps donnée et à un certain prix. Nous montrons, dans le cadre du modèle initial du *Cumulus*, que les utilisateurs peuvent « tricher » ou ne pas respecter leur contrat. Dès lors, nous proposons une modification du modèle qui permettra de prévenir de ce type de comportement. Enfin, nous déterminons les paramètres du modèle afin que le fournisseur optimise son revenu. Ces paramètres sont fixés arbitrairement dans la littérature [93].

## 2.1 Présentation du modèle du *Cumulus*

Le principe du modèle du *Cumulus* est de travailler à différentes échelles de temps. Premièrement, un contrat (SLA) est négocié entre chaque utilisateur et son fournisseur d'accès sur de longues périodes (classiquement plusieurs semaines ou mois). Ce contrat porte sur une quantité de ressource  $x$  ou consommation moyenne pendant une période fixée. L'utilisateur est facturé suivant la consommation moyenne spécifiée dans le contrat. Deuxièmement, sur de courts intervalles de temps, la consommation de chaque utilisateur est mesurée. Enfin, sur des intervalles de temps plus grands, le comportement de l'utilisateur est reporté à travers un mécanisme de *feed back*, les CPs (*Cumulus Points*). Ce mécanisme indique l'écart entre la consommation réelle (mesurée par le fournisseur) et celle spécifiée dans le contrat. Les CPs s'accumulent au cours du temps amenant à la renégociation du contrat lorsque le total dépasse un seuil  $\Theta$  (voir figure 2.1).

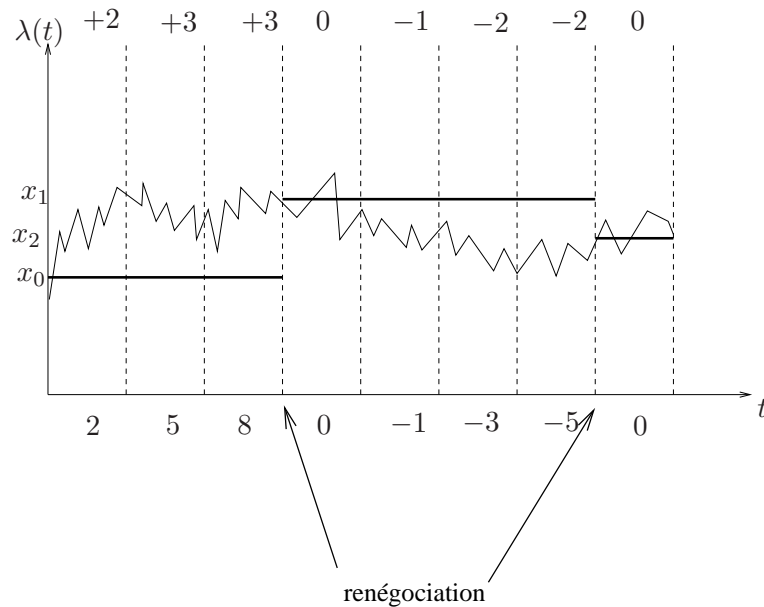


FIG. 2.1 – Illustration de l'évolution du mécanisme du *Cumulus*.

La figure 2.1 illustre l'évolution du mécanisme du *Cumulus* au cours du temps. Le total des points accumulés est représenté sous l'axe des abscisses. On remarque sur cet exemple qu'au bout de la troisième période le seuil positif  $\Theta_+ = 7$  a été atteint et donc un renégociation du contrat à la hausse est envisagé, i.e.  $x_1 > x_0$ . Ensuite, à la fin de la septième période, le seuil négatif  $\Theta_- = -5$  est atteint, et donc une renégociation à la baisse est envisagée, i.e.  $x_2 < x_1$ .

Par simplification, nous considérons par la suite comme dans [93] que les seuils de renégociation à la hausse  $\Theta_+$  et à la baisse  $\Theta_-$  du contrat sont égaux à  $\Theta$  en valeur absolue.

## 2.2 Principes de fonctionnement

Nous définissons par  $\lambda(t)$  la consommation en bande passante d'un utilisateur à l'instant  $t$  (en supposant qu'elle puisse être mesurée d'une façon assez précise via des techniques de mesures de trafic comme dans [80]) et par  $x$  la consommation moyenne requise par l'utilisateur dans son contrat avec son fournisseur de service. Le temps est découpé en intervalles réguliers  $[t_i, t_{i+1}]$  (à une échelle moyenne) avec  $i \geq 1$ , et la différence de consommation pour la période  $[t_i, t_{i+1}]$  est définie par

$$D_i = \int_{t_i}^{t_{i+1}} \lambda(t) dt - x(t_{i+1} - t_i).$$

A la fin de chaque période de mesure  $[t_i, t_{i+1}]$ , l'utilisateur se voit attribué des CPs suivant l'écart  $D_i$ . Nous posons  $CP^{\max}$  le nombre maximum de CPs qui peuvent être attribués et  $\theta_j$ , avec  $j = -CP^{\max}, \dots, CP^{\max}$ ,  $\theta_{\pm(CP^{\max}+1)} = \pm\infty$  et  $\theta_0 = 0$ , les seuils d'attribution des CPs définis de la façon suivante :

- $C_i := C(D_i)$  CPs « positifs » sont attribués si  $0 \leq \theta_{C_i} \leq D_i < \theta_{C_i+1}$ ,
- $C_i := C(D_i)$  CPs "négatifs" sont attribués si  $\theta_{C_i-1} < D_i \leq \theta_{C_i} \leq 0$ .

sur la figure 2.2 nous illustrons l'attribution des CPs avec 4 seuils positifs et 2 négatifs.

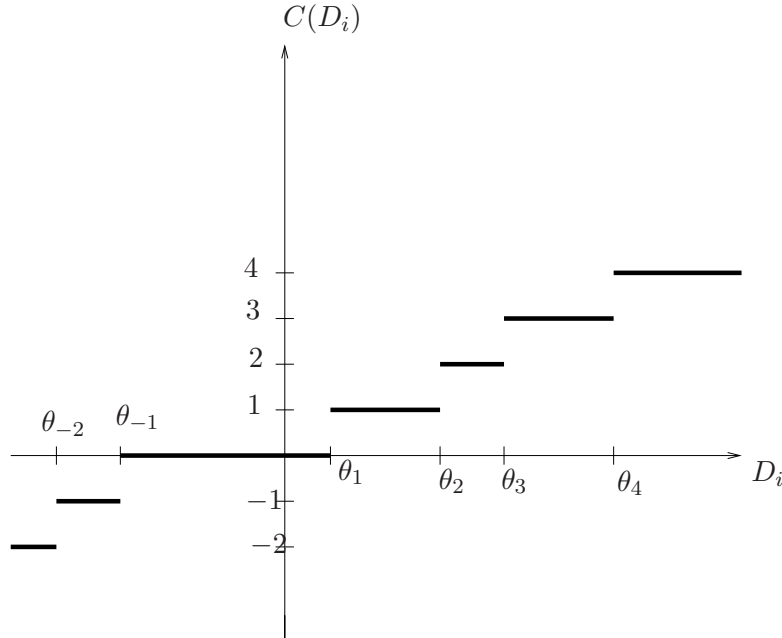


FIG. 2.2 – Attribution des CPs en fonction de  $D_i$ .

Le fournisseur décide de revoir le contrat après un nombre aléatoire de périodes  $M$  correspondant au fait que le nombre de points cumulés jusqu'à la période  $M$  dépasse

le seuil positif  $\Theta$  ou le seuil négatif  $-\Theta$ , i.e.

$$\left| \sum_{i=1}^M C_i \right| \geq \Theta.$$

- Si  $\sum_{i=1}^M C_i \geq \Theta$ , alors le contrat est négocié à la hausse. Le fournisseur et l'utilisateur vont se mettre d'accord sur une quantité de ressources  $x'$  supérieure  $x$  (renégociation à la hausse).
- Sinon, le contrat est négocié à la baisse.

Ainsi, la renégociation du contrat permet de répondre à une consommation variable à long terme si les périodes de mesure représentent plusieurs semaines par exemple.

Le contrat entre fournisseur et utilisateur porte sur la quantité de ressources allouée à l'utilisateur et sur le prix associé. Le fournisseur définit une fonction de «tarif»  $u(x)$  qui définit le prix unitaire pour une quantité de ressources  $x$  et l'utilisateur est facturé  $c(x) = xu(x)$ . La renégociation du contrat à la hausse, i.e.  $\sum_{i=1}^M C_i \geq \Theta$ , est également facturée à l'utilisateur. Ce coût financier supplémentaire est basé sur l'estimation  $d$  de la surconsommation  $D = \sum_{i=1}^M D_i$  accumulée et est facturée  $c(d)$ .

Afin d'étudier le comportement des utilisateurs face à ce nouveau modèle de tarification, nous définissons la fonction de pénalité qui dépend de la consommation totale  $\Sigma = x + d$  par

$$\Psi(x, \Sigma) = c(\Sigma) - (c(x) + c(\Sigma - x)).$$

Cette fonction ne dépend pas de la période  $i$  de mesure et donc nous ôtons l'indice pour simplifier les notations. Cette fonction de pénalité représente la différence de coût lorsqu'un utilisateur obtient  $x$  quantité de ressources dans son contrat et que sa consommation réelle mesurée est  $\Sigma$ . Afin que le modèle soit économiquement «satisfaisant», les fonctions  $u$ ,  $c$  et  $\Psi$  doivent vérifier les hypothèses suivantes [92].

**Hypothèse 1**  $\forall x > 0, u(x)$  est strictement positive et monotone décroissante (plus la quantité de ressource demandée est importante, plus le prix à l'unité est faible).

**Hypothèse 2**  $\forall x > 0, c(x) = xu(x)$  est monotone croissante (coût croissant en fonction de la quantité demandée).

**Hypothèse 3** Si  $x \neq \Sigma$ ,  $\Psi(x, \Sigma) < 0$  sinon  $\Psi(x, \Sigma) = 0$  (incitation à respecter la quantité de ressource spécifiée dans le contrat avec le fournisseur).

**Hypothèse 4** La fonction  $\Psi(x, x + d)$  est monotone décroissante en  $d$  (cette propriété assure que la pénalité croît avec l'écart à la consommation).

**Hypothèse 5**  $|\Psi(x, \Sigma)| < |\Psi(\beta x, \beta \Sigma)| \leq \beta |\Psi(x, \Sigma)|$  pour  $\beta > 1$  (passage à l'échelle : une erreur relative similaire donne une plus grande pénalité pour une demande de ressources plus importante).



Nous considérerons par exemple la fonction  $c(x) = \sqrt{x}$  pour nos applications numériques. Cette fonction vérifie l'ensemble des hypothèses précédentes.

Durant chaque période de mesure, le système mesure à certains instants, la consommation instantanée de chaque utilisateur afin d'estimer la consommation moyenne de ressource. La taille de l'échantillon, le nombre de mesures, doit être suffisamment importante pour avoir une bonne estimation de la consommation sur l'ensemble de la période. Si nous supposons que  $\lambda(t)$  est un processus stochastique en régime stationnaire, alors le théorème de la limite centrale (avec variance inconnue) pour un échantillon de  $l$  mesures nous donne un intervalle de confiance de l'estimateur  $d$  de niveau  $\alpha$ . Sa largeur est  $2\varepsilon_{\alpha,l}$  avec  $\varepsilon_{\alpha,l} = \frac{q_\alpha \hat{S}}{\sqrt{l}}$  en utilisant l'estimateur sans-biais  $\hat{S}$  de la variance du processus  $\lambda(t)$  et  $q_\alpha$ , le quantile de niveau  $(1 - \frac{\alpha}{2})$  de la loi normale centrée réduite. Ainsi, nous supposons que deux seuils voisins d'attribution des CPs sont distants d'au moins  $2\varepsilon_{\alpha,l}$ , alors le nombre de CPs attribués ne dépend pas de la façon dont a été réalisé l'échantillonnage. Nous supposons donc

$$\forall j \in \{-CP^{\max}, \dots, CP^{\max} - 1\}, \quad \theta_{j+1} - \theta_j \geq 2\varepsilon_{\alpha,l}, \quad (2.1)$$

avec  $\varepsilon_{\alpha,l} = \frac{q_\alpha \hat{S}}{\sqrt{l}}$ . Cette condition implique une distance minimale entre deux seuils voisins d'attribution des CPs. La condition sur la taille d'échantillonnage  $l$  est :

$$l \geq \frac{4\hat{S}^2 q_\alpha^2}{\min(\theta_{j+1} - \theta_j)^2}. \quad (2.2)$$

## 2.3 Définitions

A notre connaissance, il n'existe aucune hypothèse particulière dans les articles traitant du Cumulus sur la configuration des seuils d'attribution des CPs. Nous réalisons donc une étude approfondie en ce sens pour maximiser le revenu du fournisseur sous la contrainte que les utilisateurs aient intérêt à respecter leur contrat. Nous définissons pour cela en fonction des paramètres du modèle les fonctions suivantes :

- le revenu du fournisseur,
- le niveau de satisfaction d'un utilisateur.

### 2.3.1 Le revenu du fournisseur

Le revenu du fournisseur s'exprime par la différence entre ses bénéfices et ses coûts provenant des mesures de consommation par exemple. Nous décrivons ci-dessous l'ensemble des coûts du fournisseur.

- Chaque CP attribué (positif ou négatif) possède un impact financier sur l'utilisateur via une surtaxe pour chaque point et indirectement pour le fournisseur via la renégociation du contrat. Cet impact est supposé de taux  $\gamma$  linéaire en fonction du nombre de CPs attribués. Ainsi,  $\gamma C(d)$  représente l'impact de  $C(d)$  CPs, aussi bien sur le revenu du fournisseur et sur le niveau de satisfaction des utilisateurs. En effet, à long terme, une renégociation du contrat à la hausse implique un

coût à l'utilisateur ou un bénéfice au fournisseur. Quant à une renégociation à la baisse implique un manque à gagner au fournisseur et une diminution du coût à l'utilisateur.

- La taille d'échantillonnage  $l$  nécessaire pour réaliser les mesures ont également un coût. Soit  $\zeta$  le coût d'une unité d'échantillonnage alors le coût total lié aux mesures pour le fournisseur est  $\zeta l$ .
- Nous définissons également un coût provenant de l'erreur d'estimation. En effet, il semble important de pénaliser le fournisseur pour la mauvaise précision de ses mesures (car cela peut inciter des utilisateurs à rompre leur contrat). Nous supposons que ce coût est linéaire en fonction de la largeur de l'intervalle de confiance (distance minimum entre deux seuils d'attribution des CPs voisins)  $2\varepsilon_{\alpha,l} = 2\frac{\hat{S}q_{\alpha}}{\sqrt{l}}$ . Soit  $\mu$  le coût unitaire sur l'erreur, nous définissons la fonction  $F_{\text{net}}$  afin d'exprimer l'ensemble des coûts du fournisseur lié aux mesures et à l'erreur en fonction de la taille d'échantillonnage  $l$ . Cette fonction est

$$F_{\text{net}}(l) = \zeta l + 2\mu \frac{\hat{S}q_{\alpha}}{\sqrt{l}}. \quad (2.3)$$

Les bénéfices du fournisseur proviennent

- du prix  $c(x)$  provenant de la quantité de ressource  $x$  négociée dans le contrat,
- de la pénalisation  $c(d)$  liée à la surconsommation accumulée de l'utilisateur lorsque le seuil de renégociation  $\Theta$  est atteint (où  $d = \sum_{i=1}^k d_i$  est la somme des écarts de consommation depuis l'initialisation du contrat jusqu'à la  $k^{\text{eme}}$  période),
- les CPs attribués lorsque l'utilisateur consomme plus que ce qu'il demande.

Ainsi, considérant la  $k^{\text{ieme}}$  période de mesures entre deux négociations, nous pouvons définir le revenu du fournisseur pour un utilisateur.

**Définition 1** *Le revenu du fournisseur, pour la  $k^{\text{ieme}}$  période de mesures, pour un utilisateur qui a requis  $x$  unités de bande-passante et a consommé durant cette période  $x + d_k$  est*

$$R(x, \underline{d}, \underline{\theta}, k, l) = c(x) + c\left(\sum_{i=1}^k d_i\right) \mathbb{1}_{\{\sum_{i=1}^k C_i \geq \Theta\}} + \gamma C(d_k) - F_{\text{net}}(l),$$

avec  $\underline{d} = (d_1, \dots, d_k)$  et  $\underline{\theta} = (\theta_{-CP^{\max}}, \dots, \theta_{CP^{\max}})$ .

Dans cette définition,  $\underline{\theta}$  appartient à  $\mathbb{R}^{2N}$ ,  $d_k$  est la différence entre la consommation mesurée et celle spécifiée dans le contrat pour la  $k^{\text{ieme}}$  période de mesure,  $\Theta$  le nombre total de CPs accumulés nécessaires pour une renégociation du contrat.

### 2.3.2 Satisfaction de l'utilisateur

Nous introduisons une fonction de valuation  $V(x)$  permettant de mesurer les préférences de cet utilisateur. Dans notre cas, comme dans la plupart des articles liés à la tarification des réseaux, la fonction  $V(x)$  peut s'interpréter comme la quantité d'argent

qu'un utilisateur est prêt à payer afin de recevoir la quantité de ressources  $x$  [32]. Si nous considérons la  $k^{\text{ième}}$  période de mesures, l'utilité est définie comme la différence entre la satisfaction associée à sa consommation moyenne  $x + d_k$  et ses coûts. Ceux-ci comprennent la charge  $c(x)$  liée au contrat, la surtaxe  $c\left(\sum_{i=1}^k d_i\right)$  provenant de la surconsommation lorsque le seuil de renégociation  $\Theta$  est atteint et le coût  $\gamma C(c_k)$  associé aux CPs attribués pour cette période.

**Définition 2** *L'utilité (résiduelle) pour un utilisateur dont la quantité de ressources spécifiée dans le contrat est  $x$  et ayant consommé  $x + d_k$  durant la  $k^{\text{ième}}$  période s'écrit :*

$$U(x, \underline{d}, \underline{\theta}, k) = V(x + d_k) - c(x) - \gamma C(d_k) - c\left(\sum_{i=1}^k d_i\right) \mathbb{1}_{\{\sum_{i=1}^k c_i \geq \Theta\}},$$

avec  $\underline{d} = (d_1, \dots, d_k)$  et  $\underline{\theta} = (\theta_{-CP_{\max}}, \dots, \theta_{CP_{\max}})$ .

Nous remarquons que la variable aléatoire discrète  $C(d_k)$  ne dépend pas seulement de  $d_k$  mais aussi du vecteur  $\underline{\theta}$  des seuils d'attribution des CPs.

## 2.4 Problème d'incitation

Le problème d'incitation est un problème majeur dans les modèles de tarification de l'Internet comme les modèles basés sur les mécanismes d'enchères de bande-passante par exemple [101] ou les modèles multiclassés dans lesquels, les utilisateurs ont le choix de leur classe de service [81]. Ce problème d'incitation à respecter le contrat est par exemple spécifique à la téléphonie mobile où la tarification est forfaitaire et les communications hors-forfaits sont surtaxées afin de forcer les utilisateurs à respecter leur abonnement.

**Propriété 1 (Incitation à respecter son contrat)** *Le modèle du Cumulus incite les utilisateurs à respecter leur contrat si et seulement si :*

$\forall \underline{\theta}, k \quad \exists x^* > 0$  tel que

$$\arg \max_{x, d} U(x, d, \underline{\theta}, k) = (x^*, 0, \underline{\theta}, k). \quad (2.4)$$

Un utilisateur consomme la quantité de ressources  $x + d$  qui optimise son utilité et donc le mécanisme incite les utilisateurs à respecter leur contrat si la quantité demandée correspond à la consommation réelle (c'est-à-dire nous avons un écart de consommation nul, i.e.  $d = 0$ ). Nous montrons dans le théorème suivant que le modèle du Cumulus défini dans [94] (et précédemment) ne vérifie pas cette propriété.

**Théorème 1** *Considérons le mécanisme du Cumulus défini précédemment. Un utilisateur n'a pas intérêt à consommer la quantité de ressources spécifiée dans son contrat avec son fournisseur.*

**Preuve** Soit  $x$  la quantité de ressources spécifiée dans le contrat. Nous prenons le cas particulier d'un utilisateur qui consomme exactement la même quantité de ressources  $x + d_k$  sur chaque période de mesure. Ainsi,

$$U(x, d_k, \theta_k, k) = V(x + d_k) - c(x) - \gamma C(d_k) - c(kd_k) \mathbb{1}_{\{\sum_{i=1}^k C_i \geq \Theta\}}.$$

Afin de prévenir un phénomène d'hystérésis lié aux mesures, le premier seuil d'attribution des CPs  $\theta_1$  est choisi strictement positif. Supposons que  $d_k$  est compris entre 0 et  $\theta_1$ . Une condition nécessaire afin d'obtenir la relation (2.4) est :

$$\forall k, \underline{\theta}, \exists x > 0, \forall d_k \in [-x, +\infty[\setminus\{0\},$$

$$U(x - d_k, d_k, \underline{\theta}, k) < U(x, 0, \underline{\theta}, k).$$

Or, nous avons  $0 \leq d_k < \theta_1$  ce qui implique  $C_i(d_k) = 0 \forall i$  et donc

$$\sum_{i=1}^k C_i = 0.$$

Ainsi

$$U(x^* - d_k, d_k, \underline{\theta}, k) - U(x^*, 0, \underline{\theta}, k) = c(x^*) - c(x^* - d_k) > 0.$$

Ceci nous montre que l'équation (2.4) n'est pas vérifiée et donc que la condition qui implique le respect du contrat n'est pas satisfaite par ce modèle. ■

*Remarque* Il est intéressant d'observer que la preuve de ce théorème est indépendante de la fonction d'utilité.

Nous avons donc montré que la propriété n'est pas vérifiée lorsqu'un utilisateur sous-estime sa déclaration car en consommant  $x + d$  avec  $d \in [0, \theta_1[$ , il n'est pas pénalisé par l'obtention d'un point. Dans la proposition suivante, nous présentons une condition suffisante sur les seuils  $\underline{\theta}$  d'attribution des CPs contraignant l'utilisateur à limiter sa différence de consommation à  $\theta_1$ .

**Proposition 1** *Supposons que les seuils d'attribution des CPs positifs vérifient*

$$\theta_1(x) < c^{-1}(\gamma + c(x)) - x \quad (2.5)$$

et pour  $j \geq 2$ ,

$$\theta_j(x) < c^{-1}((j-1)\gamma + c(x)) - x, \quad (2.6)$$

alors  $\forall \underline{\theta}, k, x > 0, d_k \in [\theta_1, +\infty[$ , nous avons  $U(x, d_k, \underline{\theta}, k) < U(x + d_k, 0, \underline{\theta}, k)$ .

**Preuve** Soit  $x > 0$  et  $d_k \geq \theta_1$ . La différence d'utilité des utilisateurs, quand la consommation réelle est  $x + d_k$ , entre une déclaration  $x$  et une autre de  $x + d_k$  est égale à :

$$U(x, d_k, \underline{\theta}, k) - U(x + d_k, 0, \underline{\theta}, k) = c(x + d_k) - c(x) - \gamma C(d_k) - c\left(\sum_{i=1}^k d_i\right) \mathbb{1}_{\{\sum_{i=1}^k C_i \geq K\}}.$$

Une condition suffisante pour s'assurer que cette expression est négative est

$$D(d_k) = c(x + d_k) - c(x) - \gamma C(d_k) < 0 \quad \forall d_k \geq \theta_1 \quad (2.7)$$

(cette expression correspond au cas particulier sans la sur-taxe).

Supposons tout d'abord que  $d_k = \theta_1$ . Depuis l'équation (2.5), nous avons

$$D(d_1) = c(x + \theta_1) - c(x) - \gamma < 0.$$

Ensuite, comme la fonction  $D$  est strictement croissante en  $d_k$  pour  $d_k \in [\theta_{j-1}, \theta_j[$  ( $j \geq 2$ ), il est suffisant de montrer que la condition (2.7) est vérifiée lorsque  $d_k$  converge vers  $\theta_j$ . Nous avons donc

$$\lim_{\varepsilon \rightarrow 0} D(\theta_j - \varepsilon) = D(\theta_j) + \gamma = c(x + \theta_j) - c(x) - \gamma(j - 1),$$

qui donne la condition suffisante (2.6). ■

Cette proposition implique que les utilisateurs n'ont pas intérêt à dépasser leur consommation de plus que  $\theta_1$ . Plus exactement, le coût perçu pour avoir obtenu  $j - 1$  CPs est plus important que la différence de coût entre les déclarations  $x + \theta_j$  et  $x$ . Le modèle du Cumulus défini précédemment ne vérifie pas la propriété d'incitation mais il implique les utilisateurs à ne pas dépasser leur contrat d'une quantité de ressources supérieure à  $\theta_1$ , laquelle peut être choisie faible par le fournisseur. Seulement, il suffit de modifier légèrement le modèle initial et nous obtenons la propriété d'incitation. Pour cela, nous proposons dans la suite de modifier le principe de pénalisation du modèle du Cumulus précédent.

## 2.5 Le modèle du Cumulus avec pénalisation

Afin de prévenir les utilisateurs d'un non-respect de leur contrat, nous introduisons une variation au précédent mécanisme. Nous proposons de faire payer la pénalité  $c(d)$  à la fin de chaque période de mesure plutôt que d'attendre la renégociation du contrat comme spécifié dans le modèle initial. Ce nouveau mécanisme peut correspondre à un cas particulier du précédent en considérant le seuil de renégociation  $\Theta$  égal à 0. L'utilisateur est facturé dès que la quantité de ressource demandée est dépassée sur une période, c'est-à-dire dès que  $d > 0$ . Ainsi, les définitions du revenu du fournisseur et de l'utilité sont modifiées et sont maintenant indépendantes de la période. On décide alors d'ôter l'indice  $k$  d'identification de la période pour simplifier les expressions. Il faut noter que les CPs s'accumulent au cours des périodes en vue d'une renégociation du contrat.

### 2.5.1 Définitions

**Définition 3** *Pour toute période de mesure, le revenu du fournisseur pour un utilisateur ayant déclaré  $x$  quantité de ressources dans son contrat et consommant  $x + d$  est défini par*

$$R(x, d, \underline{\theta}, l) = c(x) + c(d) \mathbb{1}_{\{d > 0\}} + \gamma C(d) - F_{\text{net}}(l).$$

L'utilité d'un consommateur est

$$U(x, d, \underline{\theta}) = V(x + d) - c(x) - c(d)\mathbb{1}_{\{d>0\}} - \gamma C(d).$$

La surtaxe  $c(d)\mathbb{1}_{\{d>0\}}$  due à l'écart de consommation est facturée sur chaque période de mesure et ainsi les définitions du revenu du fournisseur et de l'utilité ne dépendent pas de la période de mesure considérée. nous montrons maintenant que ce nouveau mécanisme assure que les utilisateurs ont intérêt à respecter leur contrat.

### 2.5.2 Problème d'incitation

Nous avons modifié le modèle initial du *Cumulus* afin qu'il vérifie la propriété d'incitation à respecter de contrat décrite dans l'équation (2.4) de la propriété 1. Dans le théorème suivant, nous donnons une condition suffisante afin que l'utilisateur ne surestime pas sa consommation.

**Théorème 2** *Le modèle du Cumulus avec pénalisation satisfait la propriété d'incitation à respecter le contrat si les seuils d'attribution des CPs négatifs vérifient*

$$\forall i \geq 1 \quad \text{et} \quad x > 0, \quad c(\theta_{-i} + x) < c(x) - i\gamma.$$

*Remarque* Il est plus intéressant pour un fournisseur de connaître la demande maximum afin de faire de la prévision de trafic et du dimensionnement de son réseau. Si un utilisateur consomme moins que spécifié, cela n'a pas grande importance pour le fournisseur. Or, les conditions définies dans le théorème précédent portent uniquement sur les seuils négatifs et on peut supposer que le comportement d'un utilisateur est plutôt de dépasser sa demande.

La relation précédent peut s'exprimer également en fonction du coût  $i\gamma$  d'attribution de  $i$  CPs :

$$\forall i \geq 1 \quad \text{et} \quad x > 0, \quad i\gamma < c(x) - c(\theta_{-i} + x).$$

Le produit  $i\gamma$  représente également le bénéfice (provenant de la renégociation du contrat à la baisse et donc la diminution de sa charge) qu'obtient un utilisateur en obtenant  $i$  CPs négatifs. Ainsi, cette condition signifie que le bénéfice financier qu'obtient un utilisateur avec  $i$  CPs négatifs doit être plus petit que la différence de prix du contrat avec une quantité de ressources correspondante à  $i$  CPs négatifs. Donc un utilisateur a intérêt à renégocier son contrat plutôt qu'à obtenir des CPs négatifs. Ainsi, il n'a pas intérêt à mentir sur la valeur spécifiée dans le contrat avec son fournisseur. Nous avons besoin d'une condition suffisante afin de montrer que le nouveau mécanisme incite les utilisateurs à respecter leur contrat.

**Lemme 1** *Une condition suffisante pour que le mécanisme du cumulus avec pénalisation incite les utilisateurs à respecter leur contrat est  $\forall x > 0, \forall d \in [-x, +\infty[\setminus\{0\},$*

$$c(x + d) - c(x) - c(d)\mathbb{1}_{\{d \geq 0\}} - \gamma C(d) < 0.$$

**Preuve du lemme 1** Le mécanisme incite les utilisateurs à respecter leur contrat si et seulement si pour tout  $x > 0$  et  $d \in [-x, +\infty[\setminus\{0\}]$ , la différence entre l'utilité d'un usager ayant demandé  $x$  et ayant consommé  $x + d$ , avec un autre utilisateur ayant demandé  $x + d$  et ayant respecté son contrat, est négative. Plus formellement, cela s'écrit :

$$\forall d, x, \underline{\theta} \quad U(x, d, \underline{\theta}) - U(x + d, 0, \underline{\theta}) < 0.$$

En remplaçant l'expression de l'utilité d'un usager par son expression décrite dans la définition 3, nous obtenons directement le lemme. ■

**Preuve du théorème 2** Nous séparons les cas de surconsommation ( $d > 0$ ) et de sous consommation ( $d < 0$ ). Dans le cas de sur consommation, pour tout  $k$  et  $\underline{\theta}$ , nous avons

$$\begin{aligned} U(x, d, \underline{\theta}) - U(x + d, 0, \underline{\theta}) &= c(x + d) - c(x) - c(d) - \gamma C(d), \\ &\leq \Psi(x, x + d) < 0. \end{aligned}$$

En utilisant les hypothèses sur la fonction de pénalité  $\Psi$ , on montre que cette différence d'utilité est négative lorsque l'on considère une sur consommation d'un utilisateur, i.e.  $d > 0$ .

Considérons maintenant le cas de sous-consommation où  $d \in [-x, 0[$ . Soit  $i$  l'entier tel que  $\theta_{-i-1} < d \leq \theta_{-i}$  et donc le nombre de CPs attribués est  $C(d) = -i$ . Ainsi,  $\forall x > 0, \forall d \in [-x, +\infty[\setminus\{0\}]$ ,

$$\begin{aligned} c(x + d) - c(x) - c(d) \mathbb{1}_{\{d \geq 0\}} - \gamma C(d) &= c(x + d) - c(x) + i\gamma \\ &\leq c(x + \theta_{-i}) - c(x) + i\gamma \\ &< 0 \end{aligned}$$

en utilisant l'hypothèse sur les seuils d'attribution des CPs négatifs. Enfin le lemme 1 donne le résultat. ■

La condition suffisante décrite dans le théorème 2 porte uniquement sur les seuils d'attribution des CPs négatifs et cette condition est indépendante de la fonction d'utilité considérée. Néanmoins, cette condition n'est pas unique. En effet, une condition similaire peut provenir de la comparaison de l'utilité d'un usager ayant requis  $x$  quantités de ressources et consommant  $x + d$  avec un autre utilisateur ayant requis la même quantité mais ayant respecté son contrat. Ce qui donne

$$\forall \underline{\theta}, d, x, \quad U(x, d, \underline{\theta}) - U(x, 0, \underline{\theta}) < 0.$$

Ceci conduit aux conditions suivantes sur les seuils d'attribution des CPs négatifs  $\theta_i$  :  $\forall i > 0, \forall x > 0, \forall d \in [-x, +\infty[\setminus\{0\}]$ ,

$$V(\theta_{-i} + x) - V(x) - c(d) \mathbb{1}_{\{d > 0\}} - \gamma C(d) < 0.$$

Ces relations sont moins intéressantes car elles dépendent de la fonction d'utilité  $V$  qui est en général supposée dans la plupart des modèles de tarification de l'Internet sans

justification pratique. Cette fonction est généralement choisie suivant des propriétés mathématiques qui permettent une analyse théorique des modèles. Une autre remarque importante porte sur la dépendance des conditions avec la quantité de ressources requise dans le contrat  $x$ . Ceci aura des conséquences sur l'optimisation du revenu.

## 2.6 Optimisation du revenu du fournisseur

Nous nous intéressons maintenant à la détermination des seuils d'attribution des CPs qui maximisent le revenu du fournisseur. Les utilisateurs sont supposés rejoindre le réseau suivant un taux moyen  $\nu$  qui dépend de l'utilité décrite dans la définition 3. En supposant les conditions suffisantes du théorème 2 satisfaites, un utilisateur entre dans le réseau avec une quantité de ressources  $x$  spécifiée dans son contrat. Ainsi, l'optimisation du revenu du fournisseur pour un utilisateur, se fait pour une unique quantité de ressource donnée. Donc, l'optimisation du revenu du fournisseur est équivalent au problème d'optimisation suivant

$$\max_{\underline{\theta}} \nu(\mathbb{E}(U(x, d, \underline{\theta}))\mathbb{E}(R(x, d, \underline{\theta}, l))), \quad (2.8)$$

pour tout  $x$ , tel que

$$\forall i \geq 1 \quad \text{et} \quad x > 0, \quad c(\theta_{-i}(x) + x) < c(x) - i\gamma.$$

Nous supposons à partir de maintenant que la différence de consommation  $d$  entre la consommation réelle et celle adoptée dans le contrat est une variable aléatoire continue. Nous pouvons également supposer que la densité de  $d$  est symétrique et donc que son espérance est nulle si elle existe.

### 2.6.1 Minimisation du coût du fournisseur

La fonction  $F_{\text{net}}(l)$  donnée par l'équation (2.3) représente les coûts perçus par le fournisseur. Nous cherchons la taille de l'échantillon  $l^*$  qui minimise la fonction de coût du fournisseur sous la contrainte d'une distance minimale en seuils donnée par l'équation (2.2). La fonction  $F_{\text{net}}(l)$  est continue, dérivable et on montre après dérivation qu'elle admet un minimum global en

$$l^* = \left(\frac{\hat{S}q_\alpha\mu}{\zeta}\right)^{2/3}.$$

Nous obtenons ainsi la relation entre la taille de l'échantillonnage  $l$  et les seuils d'attribution des CPs :

- si  $l^*$  est plus grand que  $\frac{4\hat{S}^2q_\alpha^2}{\min(\theta_{i+1}-\theta_i)^2}$ , ce qui est équivalent à  $\min(\theta_{i+1} - \theta_i) \geq 2\left(\frac{\zeta}{\mu}\right)^{1/3}(\hat{S}q_\alpha)^{2/3}$ , alors le minimum de la fonction de coût est atteint en  $l = l^*$  et est égal à  $3\zeta^{1/3}(\mu\hat{S}q_\alpha)^{2/3}$ ,
- sinon le minimum est atteint en  $l = \frac{4\hat{S}^2q_\alpha^2}{\min(\theta_{i+1}-\theta_i)^2}$  et est égal à  $4\zeta\frac{\hat{S}^2q_\alpha^2}{\min(\theta_{i+1}-\theta_i)^2} + \mu\min(\theta_{i+1} - \theta_i)$ .



La fonction de coût  $F_{\text{net}}$  peut s'exprimer en fonction des seuils d'attribution des CPs par

$$F_{\text{net}}(\underline{\theta}) = L \mathbb{1}_{\{\min(\theta_{i+1}-\theta_i) \geq \kappa\}} + \left( 4\zeta \frac{\hat{S}^2 q_\alpha^2}{\min(\theta_{i+1}-\theta_i)^2} + \mu \min(\theta_{i+1}-\theta_i) \right) \mathbb{1}_{\{\min(\theta_{i+1}-\theta_i) < \kappa\}} \quad (2.9)$$

avec

$$\kappa = 2 \left( \frac{\zeta}{\mu} \right)^{1/3} (\hat{S} q_\alpha)^{2/3} \quad (2.10)$$

et

$$L = 3\zeta^{1/3} (\mu \hat{S} q_\alpha)^{2/3}.$$

Dans la suite, nous noterons  $R(x, d, \underline{\theta})$  au lieu de  $R(x, d, \underline{\theta}, l^*)$ .

### 2.6.2 Optimisation avec configurations particulières

L'optimisation du revenu du fournisseur en fonction des seuils  $\underline{\theta}$  d'attribution des CPs est difficile à traiter de façon analytique surtout à cause de la non-continuité de la fonction de coût  $F_{\text{net}}(\underline{\theta})$ . Nous considérons tout d'abord des configurations particulières des seuils où il est possible d'obtenir des résultats analytiques.

Nous traitons de façon analytique les cas particuliers suivants.

1. Nous étudions tout d'abord la configuration proposée par Reichl et al. dans [93] où les seuils sont supposés symétriques. Cette configuration peut se justifier si les utilisateurs non pas un intérêt différent à consommer plus ou moins que la quantité de ressource spécifiée dans leur contrat.
2. Ensuite, nous considérons une configuration linéaire de façon à pénaliser (respectivement récompenser) les utilisateurs proportionnellement à leur écart positif (respectivement négatif) de consommation.
3. Enfin, nous prenons en compte le cas où les seuils sont répartis de sorte que l'on ait la même probabilité d'obtenir un ou plusieurs CPs positifs et de même pour les CPs négatifs.

#### 2.6.2.1 Seuils symétriques

Nous considérons la configuration des seuils  $\underline{\theta}$  d'attribution des CPs proposé par Reichl et al. dans [93] pour laquelle nous supposons

$$\forall k, \quad \theta_k = -\theta_{-k}.$$

Ainsi le nombre moyen de CPs attribué est nul car nous avons supposé que la densité de  $d$  est symétrique. Ainsi, l'optimisation du revenu du fournisseur revient seulement à minimiser la fonction de coût  $F_{\text{net}}(\underline{\theta})$  en fonction de l'écart minimal entre deux seuils consécutifs. Donc, les seuils  $\underline{\theta}$  d'attribution des CPs sont définis tels que  $\min(\theta_{i+1} - \theta_i)$  l'écart minimum entre les seuils soit plus grand que la constante  $\kappa$  donnée dans l'équation (2.10). Ceci est équivalent à minimiser la fonction de coût  $F_{\text{net}}(\underline{\theta})$  sous la contrainte 2.2 et ainsi les seuils négatifs doivent satisfaire la condition suffisante donnée

par le théorème 2. Il faut noter également que, dans ce cas particulier, le revenu du fournisseur est indépendant du nombre de seuils ; il dépend seulement de la différence minimale entre deux seuils successifs. Ainsi le fournisseur peut considérer entre 3 et 5 seuils dans chacune des directions comme Reichl et al. dans [93].

### 2.6.2.2 Seuils linéaires

Considérons maintenant le cas des seuils  $\theta$  linéaires :

$$\forall i > 0, \quad \theta_i = i\theta_+ \quad \text{et} \quad \theta_{-i} = -i\theta_-.$$

On s'intéresse donc à déterminer  $\theta_+$  et  $\theta_-$ , les distances respectives des seuils d'attribution des CPs positifs et négatifs qui maximisent le revenu du fournisseur. Soit  $F_d$  la fonction de répartition de la variable aléatoire  $d$  et  $f_d$  sa densité. Le nombre moyen de CPs attribué à un utilisateur s'exprime par

$$\begin{aligned} \mathbb{E}(C(d)) &= \sum_{k=1}^N k (F_d((k+1)\theta_+) - F_d(k\theta_+)) \\ &\quad + \sum_{k=1}^N (-k) (F_d(-k\theta_-) - F_d(-(k+1)\theta_-)) \\ &= N - \sum_{k=1}^N F_d(k\theta_+) - \sum_{k=1}^N F_d(-k\theta_-). \end{aligned} \quad (2.11)$$

Nous considérons séparément les trois cas de figure suivants,  $\theta_+ < \theta_-$ ,  $\theta_+ > \theta_-$  et  $\theta_+ = \theta_-$ . Une comparaison numérique du maximum obtenu sur chacun des sous-domaines nous permet de déduire le maximum global.

(a)  $\theta_+ < \theta_-$  Le coût du fournisseur dépend seulement de  $\theta_+$  car  $\min(\theta_{i+1} - \theta_i) = \theta_+$ . Afin de simplifier les expressions, nous notons  $\nu$  et  $\mathbb{E}(R)$  les fonctions de débit et revenu moyen sans leurs arguments respectifs dans les équations suivantes. Une première équation est obtenue en calculant les conditions du premier ordre du problème d'optimisation.

$$\begin{aligned} \frac{\partial}{\partial \theta_-} (\nu(\cdot)\mathbb{E}(R)) &= \mathbb{E}(R)\nu' \frac{\partial}{\partial \theta_-} \mathbb{E}(U) + \nu \frac{\partial}{\partial \theta_-} \mathbb{E}(R) \\ &= \mathbb{E}(R)\nu' \gamma \sum_{k=1}^{CP_{\max}} (-k) f_d(-k\theta_-) - \nu \gamma \sum_{k=1}^{CP_{\max}} (-k) f_d(-k\theta_-), \\ &= \left( \gamma \sum_{k=1}^{CP_{\max}} (-k) f_d(-k\theta_-) \right) (\mathbb{E}(R)\nu' - \nu) \\ &= 0, \end{aligned} \quad (2.12)$$

et

$$\begin{aligned}
\frac{\partial}{\partial \theta_+} (\nu \mathbb{E}(R)) &= \mathbb{E}(R) \nu' \frac{\partial}{\partial \theta_+} \mathbb{E}(U) + \nu \frac{\partial}{\partial \theta_+} \mathbb{E}(R) \\
&= \mathbb{E}(R) \nu' \gamma \sum_{k=1}^{CP_{\max}} k f_d(k\theta_+) - \nu \gamma \sum_{k=1}^{CP_{\max}} k f_d(k\theta_+) - \nu \frac{\partial}{\partial \theta_+} F_{\text{net}}(\theta_+), \\
&= \gamma \sum_{k=1}^{CP_{\max}} k f_d(k\theta_+) (\mathbb{E}(R) \nu' - \nu) - \nu \frac{\partial}{\partial \theta_+} F_{\text{net}}(\theta_+) \quad (2.13) \\
&= 0.
\end{aligned}$$

L'équation (2.12) implique que  $\mathbb{E}(R) \nu' - \nu = 0$  car  $\gamma \sum_{k=1}^{CP_{\max}} (-k) f_d(-k\theta_-) \neq 0$ . Nous obtenons donc dans l'équation (2.13) que  $\frac{\partial}{\partial \theta_+} F_{\text{net}}(\theta_+) = 0$ . Ainsi

$$\frac{\partial}{\partial \theta_+} F_{\text{net}}(\theta_+) = \left( -8\zeta \frac{\hat{S}^2 q_\alpha^2}{\theta_+^3} + \mu \right) \mathbb{1}_{\{\theta_+ < \kappa\}},$$

donne la solution  $\theta_+^* = \kappa$  (avec  $\kappa$  défini par l'équation (2.10)). Enfin, incluant  $\theta_+^* = \kappa$  dans l'équation (2.12), on détermine  $\theta_-^*$ .

**(b)**  $\theta_+ > \theta_-$  Supposons maintenant que  $\theta_+ > \theta_-$ . Nous utilisons la même démarche que précédemment en inversant les rôles de  $\theta_+$  et  $\theta_-$ , impliquant que la fonction de coût du fournisseur dépend alors de  $\theta_-$ . Ceci donne alors la solution  $\theta_-^* = \kappa$  et  $\theta_+^*$  est obtenue en introduisant  $\theta_-^*$  dans l'une des équations obtenues via les conditions du premier ordre.

**(c)**  $\theta_+ = \theta_-$  Dans ce dernier cas de figure, les seuils sont symétriques et nous sommes alors dans le cas particulier étudié dans la section 2.6.2.1. Les solutions sont donc  $\theta_+^* = \theta_-^* = \kappa$ .

Après avoir obtenu les trois solutions ci-dessus, la détermination du maximum global s'obtient par une comparaison numérique des différents optimum. En effet, l'expression des seuils optimaux est trop complexe pour pouvoir comparer de façon analytique les optimum locaux.

### 2.6.2.3 Seuils uniformes

Soit  $CP_+$  le nombre de seuils positifs et  $CP_-$  le nombre de seuils négatifs. Nous les supposons distribués suivant les quantiles de la loi de distribution  $f_d$ , c'est-à-dire que la probabilité d'obtenir  $n$  CPs est la même pour tout  $n \leq CP_+$  (une propriété similaire est faite sur les seuils d'attribution des CPs négatifs). Nous avons donc formellement,

$$\theta_k = \begin{cases} F_d^{-1}\left(\frac{1}{2} + \frac{k}{2(CP_+ + 1)}\right) & \text{si } k > 0, \\ F_d^{-1}\left(\frac{1}{2} + \frac{k}{2(CP_- + 1)}\right) & \text{si } k < 0. \end{cases}$$

Le nombre moyen de CPs obtenu par un utilisateur est

$$\begin{aligned} \mathbb{E}(C(d)) &= \sum_{k=1}^{CP_+} k (F_d(\theta_{k+1}) - F_d(\theta_k)) + \sum_{k=1}^{CP_-} (-k) (F_d(-\theta_k) - F_d(\theta_{-k-1})) \\ &= \sum_{k=1}^{CP_+} \frac{k}{2(CP_+ + 1)} - \sum_{k=1}^{CP_-} \frac{k}{2(CP_- + 1)} \\ &= \frac{1}{4}CP_+ - \frac{1}{4}CP_-. \end{aligned}$$

Notre problème d'optimisation est donc défini en fonction des deux paramètres  $CP_+$  et  $CP_-$ . Si l'on considère par exemple que  $d$  suit une loi uniforme sur l'intervalle  $[-Z, Z]$ , les seuils sont définis par

$$\theta_k = \begin{cases} \frac{Z}{CP_+ + 1} & \text{si } k > 0, \\ -\frac{Z}{CP_- + 1} & \text{si } k < 0. \end{cases}$$

Nous étudions les conditions du premier ordre dans le but d'optimiser le revenu du fournisseur. comme dans le cas précédent avec les seuils linéaires, nous considérons plusieurs cas de figure.

**(a)**  $CP_+ > CP_-$  Dans ce cas, le plus petit écart entre les seuils est

$$\theta_1 = \frac{Z}{CP_+ + 1}. \quad (2.14)$$

tout comme dans le cas des seuils linéaires, les conditions du premier ordre sont

$$\begin{aligned} \frac{\partial}{\partial CP_-} (\nu \mathbb{E}(R)) &= \mathbb{E}(R) \nu' \frac{\partial}{\partial CP_-} \mathbb{E}(U) + \nu \frac{\partial}{\partial CP_-} \mathbb{E}(R) \\ &= \frac{\partial}{\partial CP_-} \mathbb{E}(U) (\mathbb{E}(R) \nu' - \nu) \\ &= 0 \end{aligned}$$

car on a  $\frac{\partial}{\partial CP_-} \mathbb{E}(R) = -\frac{\partial}{\partial CP_-} \mathbb{E}(U)$ .

De même  $\frac{\partial}{\partial CP_+} \mathbb{E}(R) = -\frac{\partial}{\partial CP_+} \mathbb{E}(U) - \frac{\partial}{\partial CP_+} F_{\text{net}}$ ,

$$\begin{aligned} \frac{\partial}{\partial CP_+} (\nu \mathbb{E}(R)) &= \frac{\partial}{\partial CP_+} \mathbb{E}(U) (\mathbb{E}(R) \nu' - \nu) - \nu \frac{\partial}{\partial CP_+} F_{\text{net}}(\theta_1(CP_+)) \\ &= 0. \end{aligned}$$

Comme  $\frac{\partial}{\partial CP_-} \mathbb{E}(U) = \gamma/4 \neq 0$ , ce système d'équation donne

$$\frac{\partial}{\partial CP_+} F_{\text{net}}(\theta_1(CP_+)) = 0,$$

avec l'expression de  $\theta_1$  donnée par (2.14). Ceci donne également

$$\theta_1 = \kappa, \text{ tel que } CP_+^* = \frac{1}{F_d(\kappa) - 1} - 1.$$

Afin d'obtenir  $CP_-^*$ , nous remplaçons  $CP_+^*$  par sa valeur dans la première équation donnée par les conditions du premier ordre

$$\frac{\partial}{\partial CP_-} \nu(\mathbb{E}(U(x, d, \underline{\theta}))) \mathbb{E}(R) = \frac{\gamma}{4} \mathbb{E}(R) \nu' - \frac{\gamma}{4} \nu = 0.$$

Par exemple, si nous supposons que l'arrivée des utilisateurs se fait suivant  $\nu(x) = \sqrt{x}$  et que  $d$  est uniformément distribuée sur  $[-Z, Z]$ , nous obtenons  $CP_+^* = \frac{Z}{\kappa} - 1$  et  $CP_-^*$  est la solution de

$$\frac{\gamma}{4} (A - \frac{\gamma}{4} CP_-) \frac{1}{2\sqrt{\mathbb{E}(U(x+d)) - A - L + \frac{\gamma}{4} CP_-}} - \frac{\gamma}{4} \sqrt{\mathbb{E}(U(x+d)) - A - L + \frac{\gamma}{4} CP_-} = 0,$$

avec  $A = c(x) + \mathbb{E}(c(d) \mathbb{1}_{\{d>0\}}) + \frac{\gamma}{4} CP_+^* - L$ , i.e.,

$$CP_-^* = \frac{16}{8\gamma + \gamma^2} \left( \frac{9\gamma}{4} A - 2\mathbb{E}(U(x+d)) + 2K \right).$$

**(b)  $CP_- > CP_+$**  Nous utilisons le même raisonnement que précédemment en inversant les rôles de  $CP_+$  et  $CP_-$ . Le minimum des écarts entre les seuils est alors  $\theta_{-1}$ . Nous avons donc  $CP_-^* = \frac{1}{F_d(\kappa) - 1} - 1$  et  $CP_+^*$  est obtenu en remplaçant  $CP_-^*$  par sa valeur optimale dans une des équations données par les conditions du premier ordre.

**(c)  $CP_{\max} = CP_+ = CP_-$**  Les seuils sont alors symétriques et nous retrouvons le cas particulier étudié dans la section 2.6.2.1. Depuis l'équation  $\frac{\partial}{\partial CP_{\max}} F_{\text{net}}(\theta_1(CP_{\max})) = 0$ , nous obtenons  $CP_+ = CP_- = \frac{1}{F_d(\kappa) - 1} - 1$ .

Finalement, le maximum global du revenu du fournisseur est obtenu en comparant les trois optimum locaux.

## 2.7 Applications numériques

Nous illustrons tout d'abord la propriété d'incitation à respecter le contrat et observons que le modèle initial du cumulus proposé par Reichl et al. ne la vérifie pas. Ensuite, nous montrons une application numérique d'optimisation du revenu du fournisseur pour une configuration générale des seuils  $\underline{\theta}$ . Le dernier résultat est obtenu en appliquant un algorithme d'optimisation stochastique de type recuit simulé. Finalement, nous illustrons les résultats d'optimisation obtenus dans le paragraphe précédent avec des configurations particulières des seuils  $\underline{\theta}$  d'attribution des CPs.

### 2.7.1 Compatibilité d'incitation

Nous illustrons la propriété d'incitation à respecter le contrat en représentant sur la figure 2.3 la différence d'utilité d'un usager ayant requis  $x$  quantité de ressource dans son contrat et consommant réellement  $x + d$ , avec un autre utilisateur ayant requis  $x + d$  et ayant respecté son contrat. Il est important de noter que nous présentons les résultats avec le premier modèle où  $\Theta$  est assez grand pour montrer l'importance de la condition suffisante donnée par la proposition 1.

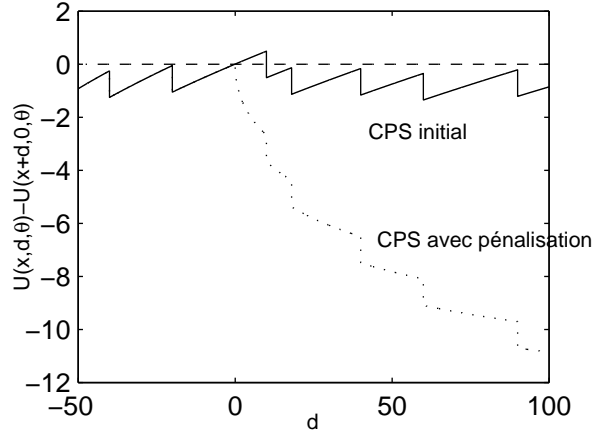


FIG. 2.3 – Différence d'utilité  $U(x, d, \underline{\theta}) - U(x + d, 0, \underline{\theta})$  en fonction de  $d$ .

Nous considérons également les seuils d'attribution des CPs

$$\underline{\theta} = [-40, -20, 0, 10, 18, 40, 60, 90],$$

de sorte que les conditions suffisantes portant sur les seuils (proposition 1) soient vérifiées pour les deux modèles. La quantité de ressource spécifiée dans le contrat est  $x = 100$  et le coût d'attribution d'un CP est  $\gamma = 1$ . Nous remarquons que pour le premier modèle, la différence d'utilité est strictement positive quand l'utilisateur dépasse son contrat d'une quantité  $d$  inférieure à  $\theta_1$ . Ceci est bien en accord avec le résultat du théorème 1 qui montre que le premier modèle ne vérifie pas la propriété d'incitation à respecter le contrat. Nous remarquons que pour le modèle adapté, cette différence d'utilité est toujours négative et ainsi ce modèle incite les utilisateurs à consommer la même quantité de ressources que celle spécifiée dans leur contrat.

### 2.7.2 Optimisation dans le cas général

L'optimisation du revenu moyen du fournisseur n'est pas évidente en parti à cause de la discontinuité de la fonction  $F_{\text{net}}(\underline{\theta})$  décrite par l'équation (2.9). De ce fait, nous utilisons une méthode d'optimisation numérique pour résoudre ce problème. Les méthodes d'optimisation globale sont classifiées comme suit :

1. les *méthodes déterministes* qui impliquent des hypothèses de dérivabilité sur la fonction  $f$  à optimiser telles que  $f$  soit continûment dérivable ou lipschitzienne.
2. Les *méthodes stochastiques* qui dépendent en partie d'événements aléatoires. Il est possible de garantir la convergence de ce type d'algorithme mais avec des contraintes fortes en pratique.

Dans notre cas d'optimisation, du fait du problème de non-dérivabilité de la fonction objectif, nous utilisons un algorithme d'optimisation stochastique de type recuit simulé.

### 2.7.2.1 Algorithme du recuit simulé

Cet algorithme (*Simulated annealing* en anglais) est une des méthodes privilégiées pour résoudre des problèmes d'optimisation combinatoire. Elle est issue d'une analogie entre le phénomène physique de refroidissement lent d'un solide en fusion et la recherche de minimums globaux d'où son appellation de *recuit simulé*. Son principe est constitué comme pour la plupart des algorithmes stochastiques, de trois étapes principales :

- une perturbation aléatoire de l'état courant du système,
- un critère d'acceptation,
- un critère d'arrêt de l'algorithme.

L'étape importante du recuit simulé est celle du critère d'acceptation qui représente l'aspect original du recuit simulé qui est de pouvoir accepter un changement d'état qui provoque une dégradation du système. Ceci permet a fortiori d'éviter les minima locaux, sous l'hypothèse de bien définir les règles de perturbation.

On remarque que dans l'algorithme du recuit simulé (algorithme 1), l'originalité se fait à l'étape où l'on applique la règle d'acceptation du nouvel état du système. Un autre phénomène important qui doit être bien maîtrisé, est le processus de refroidissement. Dans notre cas, nous utilisons une variante de cet algorithme appelé ASA pour *Adaptive Simulated Annealing*. Il est proposé par Ingber dans [57]. Cet algorithme est probablement l'algorithme de recuit simulé le plus testé dans la littérature ( pour le code, voir <http://www.ingber.com>). Ingber montre que pour garantir la convergence de l'algorithme en un temps logarithmique, la température  $T_k$  ne doit pas décroître plus vite que  $\frac{T_0}{\log(k)}$ . Le nouveau point de test s'obtient en calculant une déviation  $\Delta x$  suivant la densité suivante :

$$g(\Delta x) = (2\pi T)^{-\frac{n}{2}} \exp\left(-\frac{\|\Delta x\|^2}{2T}\right),$$

et donc le nouveau point est :  $x' = x + \Delta x$ . Le critère d'arrêt fait partie des nombreuses options du programme et dépend du nombre de nouveaux états générés, du nombre de nouveaux états acceptés et de la convergence de la température.

Pour notre application numérique, nous considérons la fonction de demande définie par

$$\nu(y) = \sqrt{y},$$

```
1: choisir, aléatoirement, une solution initiale  $x \in \mathbb{R}^D$  du système à optimiser et
   évaluer la valeur de la fonction  $f = f(x)$ ;
2: choisir une température initiale "élevée"  $T$ ;
3:  $k=1$ ;
4: perturber cette solution pour obtenir une nouvelle solution  $x' = x + \Delta x$ ;
5: calculer  $\Delta f = f(x') - f(x)$ ;
6: si  $\Delta f = f' - f \leq 0$  alors
7:   conserver la solution  $x'$ ;
8:   faire  $x \leftarrow x'$ ;
9: sinon
10:  calculer  $P = \exp(-\Delta f/T)$ ;
11:  générer un nombre aléatoire uniforme  $R$  sur  $[0, 1]$ ;
12:  si  $R \leq P$  alors
13:    accepter la nouvelle solution  $x'$ ;
14:    faire  $x \leftarrow x'$ ;
15:  sinon
16:    faire  $x \leftarrow x$ ;
17:  fin si
18: fin si
19: sauver la solution  $x$ ;
20:  $k \leftarrow k + 1$ ;
21: abaisser légèrement la température  $T \leftarrow Te^{-\log(10e^{-5})k^{1/D}}$ ;
22: si le système est figé (critère d'arrêt) alors
23:   aller au point 29;
24: sinon
25:   aller au point 3;
26: fin si
27: solution = meilleur point trouvé; fin de l'algorithme;
```

**Algorithme 1:** Algorithme du recuit simulé ASA



et nous utilisons la fonction d'utilité suivante (proposée dans [60]) :

$$V(x) = 10 \log(1 + x).$$

Nous considérons les paramètres suivants :

- quantité de ressources spécifiée dans le contrat :  $x = 100$ ,
- coût par CP attribués :  $\gamma = 100$ ,
- coût par unité d'échantillonnage :  $\zeta = 0.1$ ,
- coût de l'erreur de mesure  $\mu = 0.1$ .

Nous supposons ici que  $d$  suit une loi normale centrée réduite. Les seuils d'attribution des CPs qui donnent le revenu du fournisseur maximum sont présentés dans le tableau 2.1 pour un nombre total de seuils variant de 1 à 10.

Nombre de seuils	Position des seuils	Revenu du fournisseur
1	-2.57	54.49
2	-2.57 0.84	111.17
3	-2.57 0.92 1.85	113.17
4	-1.89 0.85 1.71 2.57	112.22
5	-2.57 -1.74 0.82 1.65 2.47	111.63
6	-2.57 -1.89 -1.22 0.67 1.34 2.01	107.31
7	-2.55 -1.9 -1.08 0.64 1.28 1.92 2.57	106.13
8	-2.43 -1.89 -0.58 0.51 1.02 1.54 2.05 2.56	98.39
9	-2.49 -1.47 -1.2 -0.77 0.33 0.87 1.36 1.61 2.51	37.54
10	-2.56 -2.11 -1.66 -1.2 -0.75 0.45 0.89 1.34 1.8 2.52	91.98

TAB. 2.1 – Seuils d'attribution des CPs maximisant le revenu du fournisseur, obtenus avec un algorithme du recuit simulé

L'algorithme d'optimisation stochastique rend la configuration des seuils donnant le meilleur revenu. De plus, il atteint son maximum avec 3 seuils dont un négatif et deux positifs. Le petit nombre de seuils est dû au coût sur les erreurs de mesure car plus le nombre de seuils est important, moins nous avons besoin de précision sur les mesures. De plus, avec seulement 1 seuil négatif, les utilisateurs n'ont pas intérêt à ne pas respecter leur contrat. Nous avons dessiné sur la figure 2.4 le revenu du fournisseur avec deux seuils (1 positif et 1 négatif) et on peut vérifier la concordance des résultats avec le tableau 2.1, ligne 2.

### 2.7.3 Optimisation avec configurations particulières

Nous illustrons les résultats analytiques obtenus dans la partie 2.6.2 pour les configurations particulières sur les seuils. Dans tous les cas de figures, nous vérifions la concordance entre les résultats analytiques et numériques.

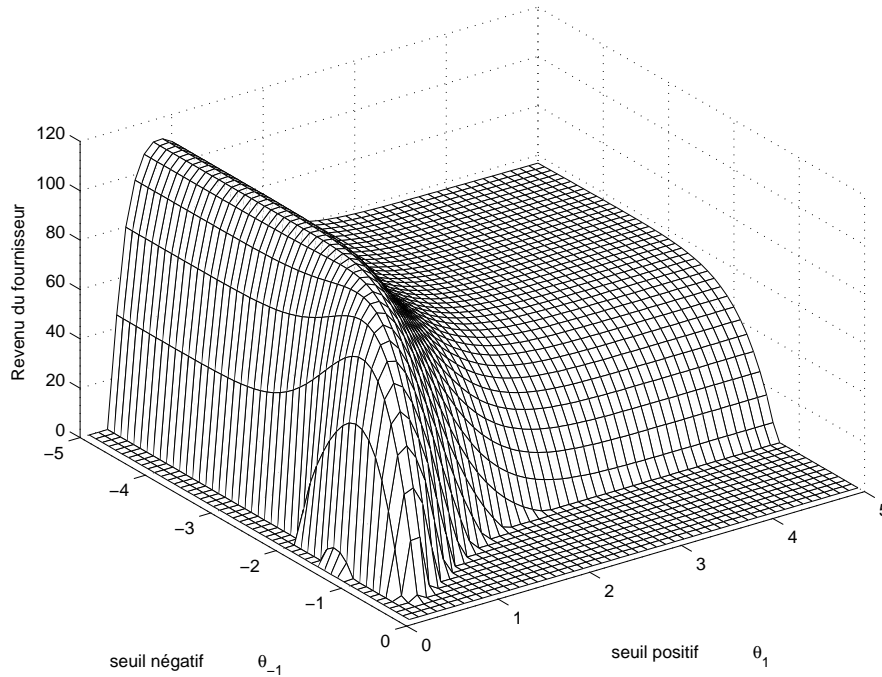


FIG. 2.4 – Revenu du fournisseur pour 1 seuil positif et 1 seuil négatif.

### 2.7.3.1 Seuils symétriques

Pour cette configuration, le revenu du fournisseur s'exprime en fonction de la distance minimal entre deux seuils consécutifs d'attribution de CPs. Sur la figure 2.5, nous avons le revenu du fournisseur en fonction de cette distance minimale pour différentes valeurs de  $x$ . La ligne verticale correspond à  $\kappa$  (équation 2.10) et nous pouvons observer qu'à partir de cette frontière, le revenu est constant. De plus, nous remarquons que le revenu du fournisseur est croissant avec la quantité de ressource  $x$  requise par l'utilisateur.

### 2.7.3.2 Seuils linéaires

Pour cet autre cas particulier, nous obtenons tout d'abord  $\kappa = 3.73$ . Nous observons sur la figure 2.6 que le revenu du fournisseur est constant lorsque  $\theta_+$  et  $\theta_-$  sont supérieurs à  $\kappa$ , car dans ce cas la fonction de coût  $F_{\text{net}}$  est indépendante des seuils. Le maximum local quand  $\theta_+ = \theta_-$  est obtenu pour  $\theta_+ = \theta_- = \kappa$  car il revient à minimiser la fonction de coût  $F_{\text{net}}$ . On peut observer que le maximum global du revenu est atteint lorsque  $\theta_+ < \theta_-$  et que l'optimum local lorsque  $\theta_- < \theta_+$  est négatif, ainsi les utilisateurs n'ont pas intérêt à entrer dans le réseau et donc le revenu est nul.

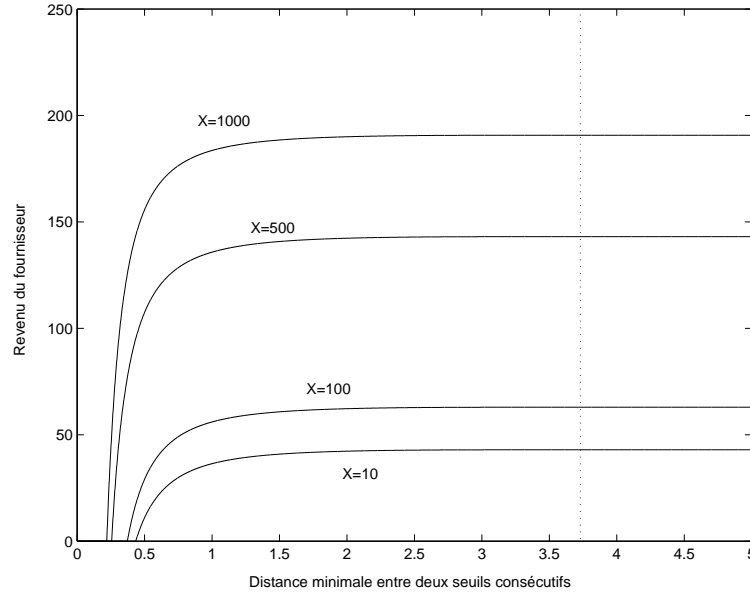


FIG. 2.5 – Revenu du fournisseur avec seuils symétriques en fonction de la distance minimale entre deux seuils consécutifs.

### 2.7.3.3 Seuils uniformes

Pour ce dernier cas particulier, nous utilisons les mêmes données que précédemment mais avec  $\zeta = 0.05$  et  $\mu = 5$ . La figure 2.7 représente le revenu du fournisseur en fonction du nombre total de seuils positifs  $CP_+$  et négatifs  $CP_-$ . Nous supposons que l'écart de consommation  $d$  suit une loi uniforme sur l'intervalle  $[-50, 50]$ . Comme dans le cas précédent, un seul optima local est valide car c'est un maximum (obtenu quand  $CP_+ > CP_-$ ) alors que dans l'autre cas, c'est un minimum local (obtenu quand  $CP_- > CP_+$ ). De plus, il faut noter que le revenu est nul lorsque l'on considère autant de seuils positifs que de seuils négatifs car le nombre moyen de CPs obtenus est zéro et donc l'utilité d'un usager ne dépend plus alors des seuils d'attribution des CP.

L'optimum obtenu est le même que celui obtenu par l'étude analytique (équations de la section 2.6.2.3) :

$$CP_+^* = 5.57 \quad \text{et} \quad CP_-^* = 2.56.$$

## 2.8 Conclusion

Notre étude a porté sur l'analyse d'un modèle de tarification adapté à une architecture de réseau à différenciation de services de type *DiffServ*, le mécanisme du Cumulus. Ce modèle est très intéressant du fait de sa simplicité et de son intégration possible dans des mécanismes de *DiffServ* comme les SLAs et le contrôle du trafic aux routeurs frontières du domaine. Ce mécanisme permet également, après une légère modification,

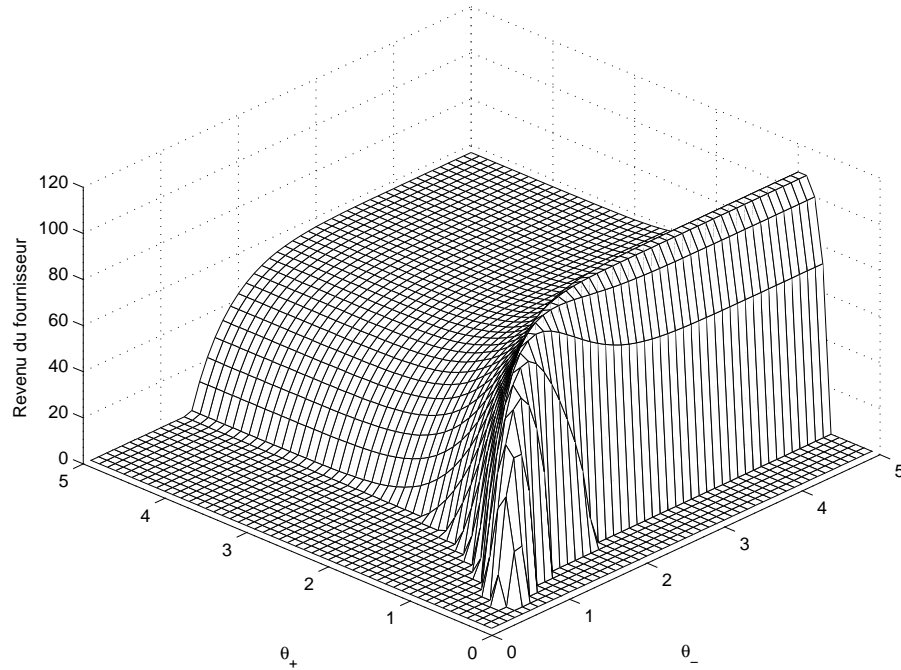


FIG. 2.6 – Revenu du fournisseur avec seuils linéaires en fonction des distances minimales  $\theta_+$  entre les seuils positifs et  $\theta_-$  entre les seuils négatifs.

de prédire la consommation moyenne des utilisateurs, ce qui est intéressant pour dimensionner le réseau. Nous avons tout d'abord montré que le modèle initial du cumulus proposé dans [94], ne répondait pas au critère d'incitation à respecter le contrat par les utilisateurs. Nous avons alors proposé une adaptation du modèle afin d'inciter ce comportement. Ensuite, nous nous sommes intéressés à l'optimisation du revenu du fournisseur en fonction d'un paramètre qui n'est pas optimisé dans la littérature : les seuils d'attribution des points. Ces paramètres sont très importants pour le mécanisme car ils permettent de contrôler les utilisateurs. Nous avons étudié de façon analytique l'optimisation du revenu du fournisseur dans les configurations particulières suivantes des seuils d'attribution des CPs :

1. *cas symétriques* : les seuils vérifient  $\theta_k = -\theta_{-k}$  pour tout  $k > 0$  ;
2. *cas linéaires* : les distances entre deux seuils consécutifs sont égales ;
3. *cas uniformes* : les seuils sont déterminés de façon que la probabilité d'obtenir  $i$  CPs soient la même pour tout  $i$ .

Dans le cas général, la non-continuité de la fonction objectif pose problème pour des méthodes d'optimisation classique basé sur le gradient. Alors, nous avons utilisé un algorithme d'optimisation stochastique basé sur le recuit simulé afin d'obtenir les seuils maximisant le revenu du fournisseur.

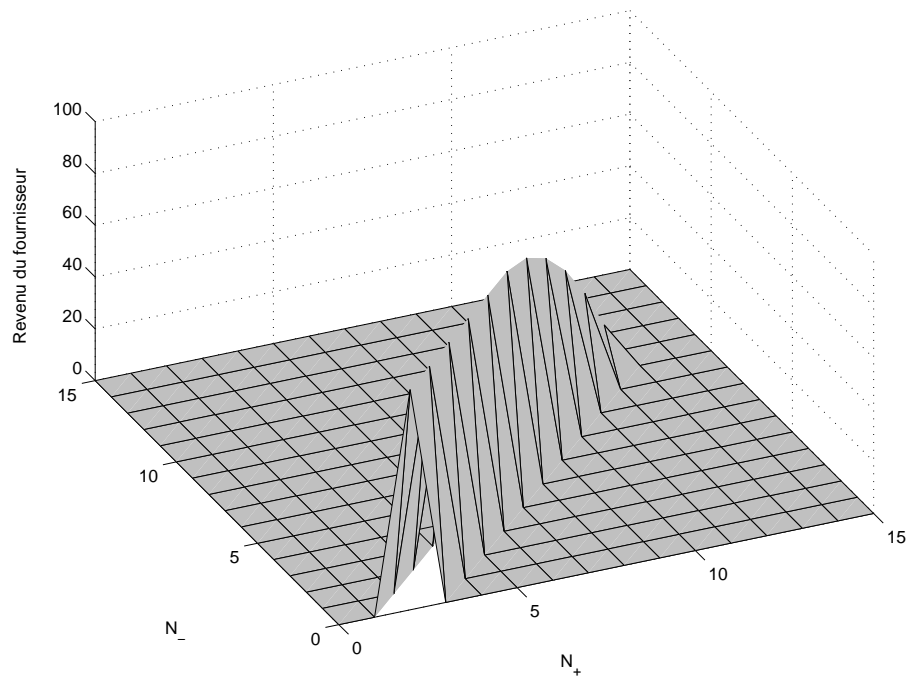


FIG. 2.7 – Revenu du fournisseur en fonction du nombre total de seuils positifs  $CP_+$  et négatifs  $CP_-$ .



## Chapitre 3

# Mécanismes d'ordonnancement et tarification

Le modèle de tarification du *Cumulus* étudié dans le chapitre précédent est prometteur car simple à implémenter dans une architecture de réseau à différenciation de services de type *DiffServ*. Son intégration dans ce type d'architecture se fait à un haut niveau, dans le contrat (SLA) entre fournisseur de service et utilisateurs. Ainsi, ce modèle ne prend pas en considération les mécanismes de différenciation mis en place dans le réseau et peut s'intégrer sans changement significatif des mécanismes de gestion du réseau. Ces mécanismes ont pourtant un impact important sur la qualité de service des classes de trafic. Le principal mécanisme mis en œuvre au sein d'un routeur de cœur d'un domaine *DiffServ* est la politique de gestion des classes de service.

Dans une architecture de réseau *DiffServ*, la différenciation de service se fait au niveau des flux. Les flux sont agrégés en une même classe de service et marqués aux routeurs frontières du domaine *DiffServ*. L'agrégation peut se faire en fonction des types d'applications et des niveaux de qualité de service requis. La différenciation du traitement se fait entre ces agrégats de trafic au niveau des routeurs internes du domaine (voir figure 3.1).

Deux mécanismes de différenciation sont mis en place dans ces routeurs internes. Nous avons tout d'abord des algorithmes de rejet sélectif comme RIO (pour *Randomly early detection with Input and Output*) [28]. Ces mécanismes opèrent à l'intérieur des classes de service en rejetant des paquets suivant le niveau de congestion. Il existe ensuite des politiques d'ordonnancement entre classes de service qui définissent celle recevant la ressource. L'implémentation de ces mécanismes pose plus ou moins de difficultés car le principe de différenciation dans *DiffServ* se fait au niveau des flux (ou agrégats de flux) alors que le réseau réel fonctionne par paquets. Pour certains algorithmes d'ordonnancement comme GPS étudié dans la suite, des émulations par paquets ont été proposés et son implémentation reste une question ouverte.

Nous proposons dans ce chapitre d'étudier l'impact de la politique d'ordonnement sur le revenu du fournisseur. On se situe alors au niveau réseau alors que le modèle du *Cumulus*, étudié dans le chapitre précédent, se situe au niveau applicatif.

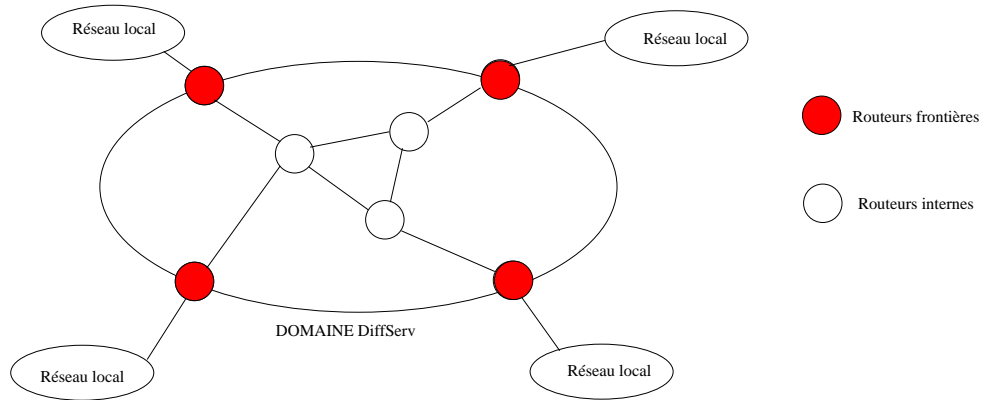


FIG. 3.1 – Routeurs frontières et internes d'un domaine *DiffServ*.

Deux politiques d'ordonnancement sont proposées pour le modèle d'architecture de réseau *DiffServ* : les priorités strictes ou les mécanismes à partage de bande passante de type WFQ (pour *Weighted Fair Queuing*). Nous considérons le mécanisme d'ordonnancement à partage de bande passante à processeur partagé généralisé noté GPS [88] (pour *Generalized Processor Sharing*). La discipline de service GPS nécessite une vision *fluide* du modèle, ce qui permet l'analyse théorique. Le mécanisme WFQ proposé dans [35] est le principal émulateur à paquets de GPS standardisé. Ce mécanisme utilise des estampilles temporelles qui indiquent pour chaque paquet son instant de départ de la file. Le calcul de cet instant de départ se fait sous l'hypothèse de la politique de service GPS. Il existe un mécanisme plus simple à files multiples qui approche le comportement de GPS, c'est la politique d'ordonnancement WRR (pour *Weighted Round Robin*) décrite sur la figure 3.2. Les classes sont servies suivant un certain ordre et le poids associé à chacune d'elles définit le nombre de paquets ou d'octets à servir durant une tournée (*round*).

Le mécanisme GPS considéré dans cette thèse est différent du modèle GPS proposé dans [30] où la capacité est partagée équitablement (à parts égales) entre tous les clients présent, mais où (contrairement à notre mécanisme GPS) la capacité dépend du nombre de clients.

Le mécanisme GPS a l'avantage d'éviter la famine des classes les moins prioritaires que l'on peut observer avec la discipline à priorités strictes. De plus, il garantit une capacité minimum à chaque classe de service. Nous comparons l'impact de cette politique de service sur le comportement des utilisateurs et optimisons le revenu du fournisseur. Nous montrons que le revenu optimal est obtenu en utilisant la politique de service avec priorités strictes. Ainsi, nous déterminons quelle est la meilleure politique d'ordonnancement pour le fournisseur entre les priorités strictes et GPS. Ce travail a été publié dans [52].

On s'intéresse ensuite à un autre mécanisme d'ordonnancement : processeur partagé de façon discriminatoire DPS (pour *Discriminatory Processor Sharing*). Le principe de fonctionnement de DPS est très proche de GPS. Dans la seconde partie de ce chapitre on



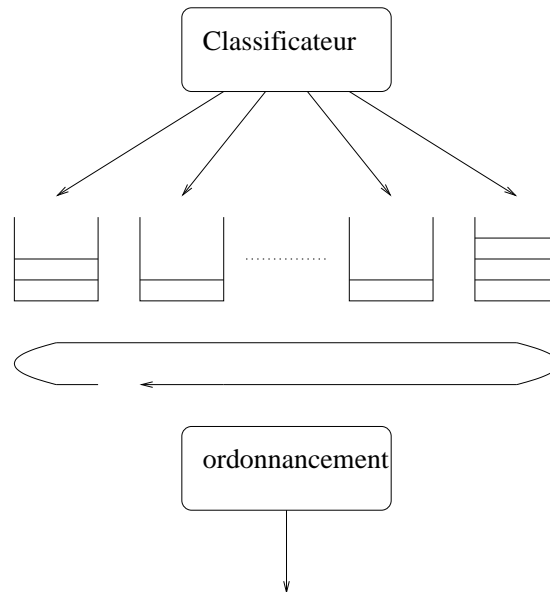


FIG. 3.2 – Mécanisme d’ordonnement WRR à files multiples.

s’intéresse donc à comparer le revenu optimal du fournisseur en utilisant les algorithmes d’ordonnement DPS et priorités strictes. Ce travail est traité dans [54] et nous montrons également, comme avec GPS, que le revenu est optimal en appliquant le mécanisme à priorités strictes. Or, dans la littérature [7, 11, 110], la modélisation de multiple trafics TCP hétérogènes à un nœud du réseau se comporte comme une file d’attente avec DPS. De plus, sur le réseau Internet, de nombreuses applications comme le mail, le transfert de données, etc, utilisent le protocole de transfert TCP dont le débit de transmission varie en fonction de la congestion. Nous pouvons appliquer alors nos résultats à un modèle de plusieurs connections TCP qui se partage un nœud du réseau. Cet autre travail a été publié dans [54].

## 3.1 Modélisation

Un modèle économique basé sur une file d’attente  $M/M/1$  multi-classes avec priorités strictes (PQ pour *Priority Queuing*) a été proposé par Mandjes dans [73]. Nous considérons le modèle avec deux types de trafic, l’un de type «voix» et l’autre de type «données», générés par une population infinie d’utilisateurs potentiels.

### 3.1.1 Modèle de trafic

Les utilisateurs sont supposés sensibles au délai moyen mis par leurs paquets pour traverser le réseau et nous considérons que les deux types de trafic se différencient par leur niveau de sensibilité. Les utilisateurs de type «voix» sont présumés plus sensibles à un long délai que les utilisateurs de type «données».

*Remarque* La dénomination des classes de service illustre le fait qu'un type de service est plus sensible à un long délai que l'autre. Il s'agit simplement de différencier les deux types de trafic.

Le système possède deux classes de service qui se différencient par leur prix d'accès. On note  $u_1$  le prix d'accès pour la première classe et  $u_2$  pour la seconde. La répartition du trafic total entre les classes de service au niveau des routeurs frontières du domaine peut se faire de deux façons.

- Dans le modèle à classes *dédiées*, le système distribue les classes de service pour chaque type de trafic. Elle peut s'appliquer par exemple pour protéger de la congestion certains type de trafic important (mis à jour des tables de routage par exemple).
- Dans le modèle à classes *ouvertes*, les utilisateurs décident de leur classe de service à un moindre coût. Ce modèle est généralement plus complexe à analyser que le précédent car on peut avoir un mélange de trafic dans la même classe de service.

### 3.1.2 Modèle des utilisateurs

Nous définissons les fonctions de valuation du service respectivement  $V_v(\cdot)$ , pour le trafic de type «voix» et  $V_d(\cdot)$ , pour le trafic de type «données». Ces fonctions dépendent du délai moyen et sont supposées strictement décroissantes. Dans ce chapitre, nous considérons les fonctions d'utilité suivantes :

$$V_d(\mathbb{E}D_d) = \frac{1}{(\mathbb{E}D_d)^{\alpha_d}} \quad \text{et} \quad V_v(\mathbb{E}D_v) = \frac{1}{(\mathbb{E}D_v)^{\alpha_v}},$$

avec  $\alpha_v$  et  $\alpha_d$  les paramètres de sensibilité face au délai. L'utilité (résiduelle) pour tout utilisateur de classe  $i \in \{v, d\}$  est alors :

$$U_i(\mathbb{E}D_i) = \frac{1}{(\mathbb{E}D_i)^{\alpha_i}} - u_i. \quad (3.1)$$

On suppose que les paramètres de sensibilité vérifient  $0 < \alpha_d < \alpha_v$ . L'idée est que les courbes d'utilité se croisent : un faible délai est préférable pour le trafic de type «voix» et inversement un grand délai est préférable pour le trafic de type «données». Il faut noter également que ce type de fonction se rapproche de la fonction de Cobb-Douglas qui est largement utilisée en analyse micro-économique [14].

Le nombre de sources ou utilisateurs pour le type «voix» est  $N_v$  et pour le type «données» est  $N_d$ . Un utilisateur de type  $i \in \{v, d\}$  rejoint le système si et seulement si son utilité  $U_i$  est positive. Ainsi, le nombre de sources (utilisateurs) augmente tant que leur satisfaction (l'utilité) est supérieure à leur coût (le prix) et chaque utilisateur agit de façon égoïste. Nous utilisons le formalisme des jeux non-coopératifs afin d'étudier ce système. Notre premier objectif est d'étudier l'existence et l'unicité d'équilibre de Nash pour ce jeu entre les types d'applications.

On considère le système, routeur interne d'un domaine *DiffServ* ou un goulot d'étranglement, modélisé par une file M/M/1 avec buffer infini et taux de service  $\mu$ . Ce modèle simple est souvent utilisé dans la littérature [6]. Les taux d'arrivées des paquets

de chaque type de trafic suivent des processus de Poisson de taux respectif  $\lambda_d$  et  $\lambda_v$  pour le type «données» et le type «voix». Le délai est alors fonction du débit  $\lambda_v N_v$  du trafic de type «voix», du débit  $\lambda_d N_d$  du trafic de type «données» et du taux de service  $\mu$ .

### 3.2 Politique de service avec priorités strictes

On considère la politique de service avec priorités strictes entre les différentes classes de service. Cette étude réalisée par Mandjes, a été publiée dans [73]. Le principal résultat porte sur la comparaison des revenus optimaux en utilisant des priorités strictes par rapport à un modèle sans différenciation de service.

**Théorème 3 (Mandjes [73])** *Dans un système composé de deux classes de service et de deux trafics de type différents, le revenu optimal, avec classes dédiées et ouvertes, en utilisant un algorithme d'ordonnancement des classes à priorités strictes, est supérieur ou égal au revenu optimum sans différenciation de service.*

En plus de promouvoir de la qualité de service pour les utilisateurs, ce résultat montre que différencier les services, permet d'engendrer un gain du revenu pour le fournisseur.

Nous nous sommes alors intéressés à la politique de service GPS, qui est plus flexible que les priorités strictes et qui est en compétition avec PQ pour être implémentée dans *DiffServ*. En effet, dans un contexte de trafic important, il peut apparaître le problème de famine des classes les moins prioritaires avec la politique à priorités strictes [112]. La politique de service GPS permet d'éviter ce problème car elle assure un taux de service minimum pour chaque classe. Ainsi, nous étudions cette politique de service dans un contexte de trafic important et comparons le revenu optimal avec celui obtenu avec priorités strictes.

### 3.3 Politique de service à processeur partagé généralisé (GPS)

Considérons le système modélisé par une file M/M/1 avec une politique de service GPS et  $K$  classes de trafic. Chacune des classes  $i \in \{1, \dots, K\}$  possède une proportion  $\gamma_i$  du taux de service total  $\mu$ , c'est-à-dire que l'on a  $\sum_{i=1}^K \gamma_i = 1$ . A l'intérieur de chaque file, le mécanisme d'ordonnancement des paquets est Premier Arrivé Premier Servi (PAPS), voir figure 3.3. Lorsqu'une file est vide, son taux de service est partagé entre les autres files non-vides équitablement et ainsi cette discipline conserve le travail (*work conserving*). Le mécanisme d'ordonnancement GPS assure donc à la classe de service  $i$  un débit minimum  $\gamma_i \mu$  à tout instant.

En effet, si l'on note  $\mathcal{B}$  l'ensemble des files non-vides, la file  $i \in \mathcal{B}$  admet un taux de service défini par :

$$\frac{\gamma_i}{\sum_{j \in \mathcal{B}} \gamma_j} \mu \geq \gamma_i \mu.$$

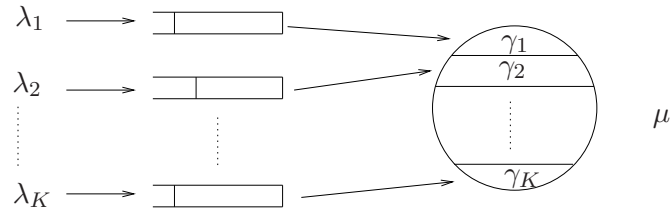


FIG. 3.3 – Le mécanisme d'ordonnancement GPS avec  $K$  classes de service.

Ce mécanisme d'ordonnancement est donc intéressant pour les applications temps-réel ou multimédia qui nécessitent des délais faibles pour avoir une bonne qualité de service. Le délai en régime stationnaire pour ce type de discipline n'est pas connu. De plus, les probabilités stationnaires ne sont pas à forme produit [22] car le taux de service de chaque file dépend de l'état du système. Cependant, si l'on se place dans le contexte de trafic important, la probabilité d'avoir une classe vide est très faible. On peut alors approché un modèle de file M/M/1 avec la politique de service GPS par  $K$  files M/M/1 indépendantes ayant chacune un taux de service  $\gamma_i\mu$  (voir la figure 3.4). Nous supposons donc que  $\mathcal{B} = \{1, \dots, K\}$ .

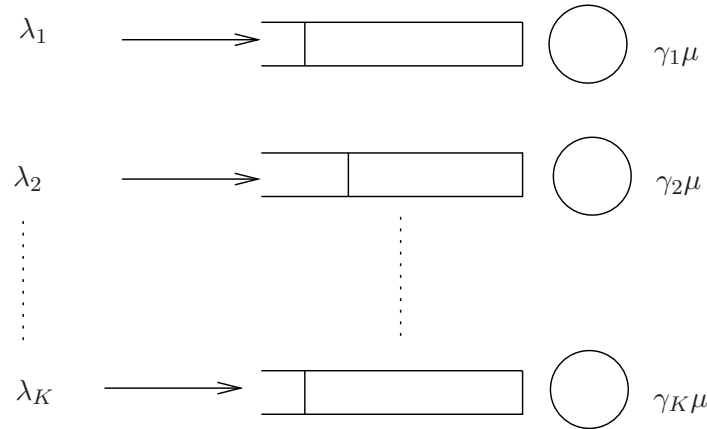


FIG. 3.4 – Le mécanisme d'ordonnancement GPS approché par un système avec  $K$  files M/M/1 indépendantes.

Si l'on considère un prix différent pour chacune des classes, cette séparation logique en  $K$  files d'attente correspond au modèle de tarification du *Paris Metro Pricing* (PMP) [86] dont les performances ont été étudiées dans [47, 97].

Nous allons donc travailler sur ce modèle approché de GPS composé de multiple files d'attente M/M/1 logiques indépendantes. Dans la suite, on ne fera pas la distinction dans les notations entre GPS et son approximation.

On note dans la suite  $\Pi_D^*$  le revenu optimal pour le cas avec classes dédiées et service avec priorités strictes,  $\Pi_O^*$  le revenu optimal pour le cas avec classes ouvertes et service avec priorités strictes,  $\Pi_{PAPS}^*$  le revenu optimal pour le modèle sans différenciation et enfin  $\Pi_{GPS}^*$  le revenu optimal pour le cas avec GPS.

### 3.3.1 Optimisation du revenu du fournisseur

Soit  $0 < \gamma < 1$  la proportion de bande-passante allouée à la classe  $i = 2$ . Ainsi, le taux de service de la classe 1 est  $(1 - \gamma)\mu$  et de la classe 2 est  $\gamma\mu$ . Nous considérons tout d'abord le cas des classes dédiées pour lequel chaque type de trafic est servi par une classe de service différente.

#### 3.3.1.1 Classes dédiées

Nous supposons que la première classe de service (file d'attente  $i = 1$ ) est dédiée au trafic de type «voix» et que la seconde classe de service (file d'attente  $i = 2$ ) est dédiée au trafic de type «données». Le trafic de type «voix» ne correspond pas à un trafic réellement multimédia, nous choisissons cette dénomination afin de le différencier avec le trafic de type «données» qui est supposé moins sensible au niveau de qualité de service. Nous notons  $u_1$  le prix pour la première classe dédiée au trafic de type «voix» et  $u_2$  le prix pour la seconde dédiée au trafic de type «données». Le délai moyen pour chacune des classes est donc

$$\mathbb{E}D_1 = \frac{1}{(1 - \gamma)\mu - \lambda_v N_v} \quad \text{et} \quad \mathbb{E}D_2 = \frac{1}{\gamma\mu - \lambda_d N_d}.$$

Le nombre  $N_v$  d'utilisateurs de type «voix» à l'équilibre du système dépend du prix  $u_1$ , du taux d'arrivée des paquets  $\lambda_v$  et du taux de service  $(1 - \gamma)\mu$ . En effet, les utilisateurs joignent le système tant que leur utilité (résiduelle) est positive; donc, à l'équilibre, le nombre d'utilisateurs  $N_i$  pour la classe  $i$  avec  $i \in \{v, d\}$  vérifie :

$$U_i(N_i) = 0.$$

**Propriété 2** *L'équilibre du système existe et les nombres d'utilisateurs  $N_v$  de type voix et  $N_d$  de type données sont pour cet équilibre :*

$$N_v = \frac{(1 - \gamma)\mu - \sqrt[\alpha]{u_1}}{\lambda_v} \quad (3.2)$$

si  $u_1 < ((1 - \gamma)\mu)^{\alpha_v}$ , 0 sinon et

$$N_d = \frac{\gamma\mu - \sqrt[\alpha]{u_2}}{\lambda_d} \quad (3.3)$$

si  $u_2 < (\gamma\mu)^{\alpha_d}$  et 0 sinon.

**Preuve** D'après l'expression des délais moyens, l'utilité  $U_i$  pour le trafic de type  $i \in \{v, d\}$  décrite par l'équation (3.1) s'exprime en fonction de  $N_i$  par :

$$U_i(N_i) = (\gamma_i\mu - \lambda_i N_i)^{\alpha_i} - u_i,$$

avec  $\gamma_v = 1 - \gamma$  et  $\gamma_d = \gamma$ . A l'équilibre, le nombre d'utilisateurs  $N_i$  pour la classe  $i$  avec  $i \in \{v, d\}$  vérifie  $U_i(N_i) = 0$ , ce qui est équivalent à :

$$N_i = \frac{\gamma_i\mu - \sqrt[\alpha]{u_i}}{\lambda_i}.$$

Dans le cas où  $u_i < (\gamma_i \mu)^{\alpha_i}$ , le prix d'accès est trop important car les utilisateurs ont une utilité résiduelle négative, donc le nombre d'utilisateurs  $N_i$  est zéro. ■

Le revenu du système est défini comme la somme du revenu obtenu avec les deux classes de service. Le revenu par unité de temps pour la classe de service  $i$  est  $\lambda_i N_i u_i$ . Ainsi, le revenu total  $R_{GPS}^{(\gamma)}$  du fournisseur est :

$$\begin{aligned} R_{GPS}^{(\gamma)}(u_1, u_2) &= \lambda_1 N_1 u_1 + \lambda_2 N_2 u_2 \\ &= \lambda_v N_v u_1 + \lambda_d N_d u_2. \end{aligned}$$

Lorsque les deux types de trafic sont présents dans le système, i.e.  $(u_1, u_2) \in D_0 = [0, ((1-\gamma)\mu)^{\alpha_v}] \times [0, (\gamma\mu)^{\alpha_d}]$  ( $N_d, N_v \geq 0$ ), le revenu s'exprime à partir des équations (3.2) et (3.3) par :

$$R_{GPS}^{(\gamma)}(u_1, u_2) = (1-\gamma)\mu u_1 - u_1^{1+1/\alpha_v} + \gamma\mu u_2 - u_2^{1+1/\alpha_d}. \quad (3.4)$$

Le système cherche à optimiser son revenu en fonction des prix et du paramètre de proportion de service  $\gamma$ . Ce problème d'optimisation s'écrit :

$$\max_{u_1 > 0, u_2 > 0, \gamma \in [0, 1]} R_{GPS}^{(\gamma)}(u_1, u_2).$$

On s'intéresse tout d'abord à l'optimisation du revenu du fournisseur lorsque le paramètre  $\gamma$  est fixé. Et les prix qui optimisent le revenu du fournisseur sur ce sous-espace sont donnés dans le théorème suivant.

**Théorème 4** *Le paramètre de partage de service  $\gamma$  étant fixé, les prix qui maximisent le revenu  $R_{GPS}^{(\gamma)}$  sont donnés par :*

$$u_1^* = \left( \frac{(1-\gamma)\mu}{1 + \frac{1}{\alpha_v}} \right)^{\alpha_v} \quad \text{et} \quad u_2^* = \left( \frac{\gamma\mu}{1 + \frac{1}{\alpha_d}} \right)^{\alpha_d}. \quad (3.5)$$

**Preuve** Le revenu du fournisseur  $R_{GPS}^{(\gamma)}(u_1, u_2)$  donné par l'équation (3.4) est une fonction continue et dérivable sur le domaine  $D_0$ . Les conditions du premier ordre pour l'optimisation du revenu  $R_{GPS}^{(\gamma)}$  sont :

$$\frac{\partial R_{GPS}^{(\gamma)}}{\partial u_1}(u_1, u_2) = (1-\gamma)\mu - \left(1 + \frac{1}{\alpha_v}\right) u_1^{1/\alpha_v} = 0$$

et

$$\frac{\partial R_{GPS}^{(\gamma)}}{\partial u_2}(u_1, u_2) = \gamma\mu - \left(1 + \frac{1}{\alpha_d}\right) u_2^{1/\alpha_d} = 0.$$

Ainsi, les prix  $(u_1^*, u_2^*)$  satisfaisant le système des conditions du premier ordre sont :

$$(u_1^*, u_2^*) = \left( \left( \frac{(1-\gamma)\mu}{1 + \frac{1}{\alpha_v}} \right)^{\alpha_v}, \left( \frac{\gamma\mu}{1 + \frac{1}{\alpha_d}} \right)^{\alpha_d} \right).$$

Le couple  $(u_1^*, u_2^*)$  est à l'intérieur de  $D_0$  et nous avons

$$\frac{\partial R_{\text{GPS}}^{(\gamma)}}{\partial u_1}(u_1, u_2) > 0 \Leftrightarrow u_1 < u_1^*,$$

et

$$\frac{\partial R_{\text{GPS}}^{(\gamma)}}{\partial u_2}(u_1, u_2) > 0 \Leftrightarrow u_2 < u_2^*,$$

donc le point  $(u_1^*, u_2^*)$  est un maximum local. Finalement, on peut vérifier via l'équation (3.4) que, à la frontière du domaine  $D_0$ , le revenu  $R_{\text{GPS}}^{(\gamma)}$  est plus faible car soit un prix ou le nombre d'utilisateurs est nul pour au moins une des files. En effet, sur les bords inférieurs nous avons :

$$R_{\text{GPS}}^{(\gamma)}(u_1^*, 0) < R_{\text{GPS}}^{(\gamma)}(u_1^*, u_2^*) \quad \text{et} \quad R_{\text{GPS}}^{(\gamma)}(0, u_2^*) < R_{\text{GPS}}^{(\gamma)}(u_1^*, u_2^*).$$

Sur les bords supérieurs, on a les inégalités suivantes :

$$R_{\text{GPS}}^{(\gamma)}(u_1^*, (\gamma\mu)^{\alpha_d}) = \lambda_1 N_1 u_1^* < \lambda_1 N_1 u_1^* + \lambda_2 N_2 u_2^* = R_{\text{GPS}}^{(\gamma)}(u_1^*, u_2^*),$$

et

$$R_{\text{GPS}}^{(\gamma)}((1 - \gamma\mu)^{\alpha_v}, u_2^*) = \lambda_2 N_2 u_2^* < \lambda_1 N_1 u_1^* + \lambda_2 N_2 u_2^* = R_{\text{GPS}}^{(\gamma)}(u_1^*, u_2^*).$$

Donc,  $(u_1^*, u_2^*)$  est le maximum global du revenu. ■

L'expression du revenu optimal du fournisseur donnée par l'équation (3.4) peut être reformulée seulement en fonction du paramètre  $\gamma$  de partage de ressource par :

$$R_{\text{GPS}}^*(\gamma) = (1 - \gamma)^{1+\alpha_v} A(\alpha_v) + \gamma^{1+\alpha_d} A(\alpha_d), \quad (3.6)$$

avec  $A(x) = \mu^{1+x} \left(\frac{x}{1+x}\right)^x \frac{1}{1+x}$ . On s'intéresse donc maintenant à l'optimisation du paramètre de partage de ressource  $\gamma$  entre les classes de service.

**Théorème 5** *Le revenu maximal  $R_{\text{GPS}}^*$  du modèle approché de GPS avec classes dédiées est donné par :*

$$R_{\text{GPS}}^* = \max(A(\alpha_v), A(\alpha_d)).$$

**Preuve** En considérant les prix optimaux donnés dans le théorème 4, le revenu du fournisseur en fonction de  $\gamma$  s'écrit :

$$R_{\text{GPS}}^*(\gamma) = (1 - \gamma)^{1+\alpha_v} A(\alpha_v) + \gamma^{1+\alpha_d} A(\alpha_d), \quad (3.7)$$

avec  $A(x) = \mu^{1+x} \left(\frac{x}{1+x}\right)^x \frac{1}{1+x}$ . Sa dérivée seconde est égale à :

$$R_{\text{GPS}}^{*''}(\gamma) = (1 + \alpha_v)\alpha_v(1 - \gamma)^{\alpha_v-1} A(\alpha_v) + (1 + \alpha_d)\alpha_d\gamma^{\alpha_d-1} A(\alpha_d) \geq 0,$$

De plus,

$$R_{\text{GPS}}^{*'}(0) = -(1 + \alpha_v)A(\alpha_v) < 0 \quad \text{et} \quad R_{\text{GPS}}^{*'}(1) = (1 + \alpha_d)A(\alpha_d) > 0.$$

Comme le revenu  $R_{\text{GPS}}^*(\gamma)$  est continu et strictement convexe ; il atteint son maximum en  $\gamma = 0$  ou  $\gamma = 1$ . Un unique type de trafic est donc servi lorsque le système optimise son revenu, c'est-à-dire :

$$\max_{\gamma} R_{\text{GPS}}^*(\gamma) = \max(R_{\text{GPS}}^*(0), R_{\text{GPS}}^*(1)) = \max(A(\alpha_v), A(\alpha_d)).$$

■

Nous avons donc montré que lorsque le système détermine les prix et le paramètre de partage de ressources qui maximisent son revenu, un unique type de trafic est présent dans le système. On s'intéresse donc à savoir quel type de trafic est présent : «voix» ou «données». La proposition suivante donne des conditions nécessaires sur les paramètres de sensibilités afin de connaître le type de trafic présent dans le système pour le revenu optimal.

**Proposition 2** *Lorsque le taux de service est plus petit que 1, seul le trafic de type «données» est traité par le système. Sinon, lorsque  $\mu > 1$ ,*

- Si  $\alpha_d > \frac{1}{\mu-1}$ , seul le trafic de type «voix» est traité.
- Si  $\alpha_v < \frac{1}{\mu-1}$ , seul le trafic de type «données» est traité.

**Preuve** Pour savoir quel type de trafic est choisi par le système, il suffit de comparer  $A(\alpha_v)$  obtenu avec  $\gamma = 0$  (trafic de type «voix») et  $A(\alpha_d)$  obtenu avec  $\gamma = 1$  (trafic de type «données»). Sa première dérivée s'écrit :

$$A'(x) = \left(\frac{\mu}{1+x}\right)^{1+x} x^x \left(\ln\left(\frac{x}{1+x}\right) + \ln \mu\right).$$

Cette expression est négative si et seulement si :

$$\ln\left(\frac{x}{1+x}\right) + \ln \mu < 0 \quad \Leftrightarrow \quad (\mu - 1)x < 1.$$

Si  $0 < \mu \leq 1$  alors  $(\mu - 1)x \leq 0, \forall x > 0$ , ainsi la fonction  $A(\cdot)$  est décroissante sur  $[0, +\infty[$ . Comme  $\alpha_v > \alpha_d$  par hypothèse, nous avons alors  $A(\alpha_v) < A(\alpha_d)$ , et ainsi le maximum est atteint en  $\gamma = 1$ ; en d'autres mots, seul le trafic de type «données» est traité lorsque  $0 < \mu \leq 1$ .

Lorsque  $\mu > 1$ ,  $A(x)$  atteint son minimum en  $x = 1/(\mu - 1)$ . Selon les valeurs de  $\mu$ ,  $\alpha_v$  et  $\alpha_d$ , on en déduit les trois cas suivants :

- Si  $\alpha_d > \frac{1}{\mu-1}$ , alors le revenu est maximisé en prenant  $\gamma = 0$  (i.e., seul le trafic de type «voix» est accepté).
- Si  $\alpha_v < \frac{1}{\mu-1}$ , alors le revenu est maximisé en prenant  $\gamma = 1$  (i.e., seul le trafic de type «données» est traité).
- Si  $\alpha_d < \frac{1}{\mu-1} < \alpha_v$ , alors il faut comparer numériquement les valeurs de  $A(\alpha_d)$  et  $A(\alpha_v)$  afin de trouver le revenu maximal.

■



Nous avons donc déterminé les prix  $(u_1^*, u_2^*)$  et le partage de ressources  $\gamma^*$  donnant le revenu maximal pour le fournisseur dans le cas où le système choisit les classes de service pour chaque type de trafic. Dans le cas optimal, une unique classe de service est traitée.

On suppose maintenant que le fournisseur propose plusieurs classes de service. Les utilisateurs ont le choix de leur classe de service. Ce choix peut être motivé par des considérations économiques ou en fonction de la qualité de service perçue. Dans la suite, nous nous intéressons à ce cas ouvert.

### 3.3.1.2 Classes ouvertes

Nous supposons dans cette section que les utilisateurs ont le choix de leur classe de service. Il est donc possible d'avoir des paquets de type «voix» et de type «données» dans la même classe de service. Nous conservons les notations de la section précédente avec les deux types de trafic «voix», «données» mais pour deux classes de service notées 1 et 2. Nous déterminons tout d'abord la nature du trafic dans chacune des classes en fonction des prix. La proposition suivante montre qu'à l'équilibre, seul un unique type de paquets est présent dans chacune des files indépendamment de l'autre file. Ensuite, nous déterminons les prix et paramètre de partage qui maximisent le revenu du fournisseur.

**Proposition 3** *Pour chaque classe de service  $i$  avec  $(i \in \{1, 2\})$ , nous avons les propriétés suivantes :*

- le trafic de type «données» est présent dans la classe de service  $i$  si et seulement si  $u_i < 1$ .
- le trafic de type «voix» est présent dans la classe de service  $i$  si et seulement si  $u_i > 1$ .

**Preuve** Considérons la  $i^{\text{eme}}$  classe de service avec  $i \in \{1, 2\}$  et  $\gamma_i$  l'allocation de ressource pour cette classe. On suppose que  $N_d$  utilisateurs de type «données» sont présents dans cette file, avec :

$$N_d(u_i) = \frac{\gamma_i \mu - \sqrt[\alpha_d]{u_i}}{\lambda_d},$$

si  $u_i < (\gamma_i \mu)^{\alpha_d}$ . Un utilisateur de type «voix» rejoindra cette file si et seulement si son utilité résiduelle  $U_v$  est positive. Celle-ci s'exprime par l'équation suivante :

$$U_v(u_i) = (\gamma_i \mu - \lambda_d N_d(u_i))^{\alpha_v} - u_i = u_i^{\alpha_v / \alpha_d} - u_i.$$

Sa fonction d'utilité est donc strictement positive si et seulement si  $u_i > 1$ . Finalement, si  $u_i > 1$ , les utilisateurs de type «voix» joignent le système et les utilisateurs de type «données» vont le quitter car leur utilité résiduelle devient négative. Nous avons donc finalement que des utilisateurs de type «voix» dans la classe de service. Inversement, si il y a  $N_v(u_i)$  utilisateurs de type «voix» dans la file, un utilisateur infinitésimal de type «données» entrera dans le réseau si et seulement si  $u_i < 1$ . ■

Les prix induisent donc différentes configurations du trafic dans le système et le revenu  $R_{GPS}^{(\gamma)}$  s'exprime alors de différentes manières. Pour  $\gamma$  fixé, le domaine d'optimisation des prix se décompose en quatre sous-domaines (voir figure 3.5) suivant la répartition des types de trafic dans les classes de service. Nous définissons les sous-domaines suivant en fonction du couple des prix  $(u_1, u_2)$  :

- Le sous-domaine  $V_O$  où seul le trafic de type «voix» est dans le système (files 1 et 2). D'après la proposition 3, nous avons donc  $u_1 > 1$  et  $u_2 > 1$ . De plus, d'après l'expression 3.2, le nombre d'utilisateurs de type «voix» dans la classe  $i$  est positif si et seulement si  $u_i < (\gamma_i \mu)^{\alpha_v}$ , d'où

$$V_O = [1, ((1 - \gamma)\mu)^{\alpha_v}] \times [1, (\gamma\mu)^{\alpha_v}].$$

- Le sous-domaine  $D_O$  où seul le trafic de type «données» est dans le système. D'après la proposition 3, nous avons donc  $u_1 < 1$  et  $u_2 < 1$ . Le nombre d'utilisateurs de type «données» dans la classe  $i$  est positif si et seulement si  $u_i < (\gamma_i \mu)^{\alpha_d}$  d'où

$$D_O = [0, \min(1, ((1 - \gamma)\mu)^{\alpha_d})] \times [0, \min(1, (\gamma\mu)^{\alpha_d})].$$

- Le sous-domaine  $M_O^{(1)}$  où le trafic de type «voix» est dans la classe 1 et le trafic de type «données» dans la classe 2. D'après la proposition 3, nous avons donc  $u_1 > 1$  et  $u_2 < 1$ . Ce sous-domaine correspond à l'encadrement suivant :

$$M_O^{(1)} = [1, ((1 - \gamma)\mu)^{\alpha_v}] \times [0, \min(1, (\gamma\mu)^{\alpha_d})].$$

- Enfin, le sous-domaine  $M_O^{(2)}$  où le trafic de type «voix» est dans la classe 2 et le trafic de type «données» dans la classe 1. D'après la proposition 3, nous avons donc  $u_1 < 1$  et  $u_2 > 1$ , ce qui correspond à l'intervalle :

$$M_O^{(2)} = [0, \min(1, ((1 - \gamma)\mu)^{\alpha_d})] \times [1, (\gamma\mu)^{\alpha_v}].$$

Afin de trouver les prix optimaux en considérant le partage de ressource  $\gamma$  fixé, nous étudions le revenu optimal du fournisseur sur chacun des sous-domaines puis nous les comparons. Ensuite nous déterminons le paramètre optimal  $\gamma$  sur chacun des sous-domaines  $V_O$ ,  $D_O$ ,  $M_O^{(1)}$  et  $M_O^{(2)}$ .

Sur le sous-domaine  $V_O$ , seul le trafic de type «voix» entre dans le réseau, et ce trafic homogène est distribué entre les deux files. Nous supposons que  $((1 - \gamma)\mu)^{\alpha_v} \geq 1$  et  $(\gamma\mu)^{\alpha_v} \geq 1$  de sorte que le sous-domaine  $V_O$  ne soit pas vide. Ceci est vrai si et seulement si :

$$0 \leq \frac{1}{\mu} \leq \gamma \leq 1 - \frac{1}{\mu} \leq 1. \quad (3.8)$$

**Lemme 2** *Les prix optimaux sur le sous-espace  $V_O$  sont :*

$$u_1^* = \max \left( 1, \left( \frac{(1 - \gamma)\mu}{1 + 1/\alpha_v} \right)^{\alpha_v} \right) \quad \text{et} \quad u_2^* = \max \left( 1, \left( \frac{\gamma\mu}{1 + 1/\alpha_v} \right)^{\alpha_v} \right).$$

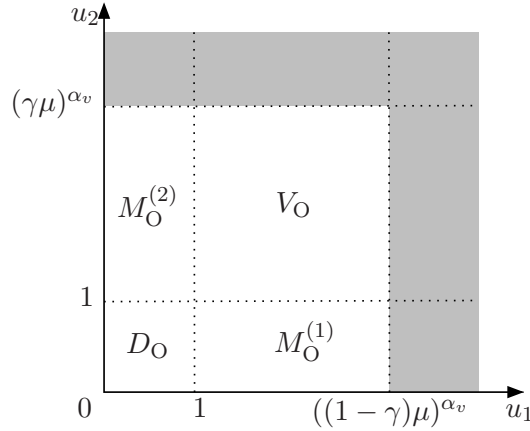


FIG. 3.5 – Répartition possible du trafic suivant les prix dans chacun des classes de service.

**Preuve** Sur le sous-domaine  $V_O$ , seul le trafic de type «voix» est présent dans le système. Donc, les nombres d'utilisateurs  $N_1(u_1)$  et  $N_2(u_2)$  dans les files 1 et 2 sont donnés par :

$$N_1(u_1) = \frac{(1-\gamma)\mu - \sqrt[\alpha_v]{u_1}}{\lambda_v} \quad \text{et} \quad N_2(u_2) = \frac{\gamma\mu - \sqrt[\alpha_v]{u_2}}{\lambda_v}.$$

Le revenu du fournisseur s'écrit donc :

$$\begin{aligned} R_{\text{GPS}}^{(\gamma)}(u_1, u_2) &= \lambda_v N_1(u_1) u_1 + \lambda_v N_2(u_2) u_2 \\ &= (1-\gamma)\mu u_1 - u_1^{1+1/\alpha_v} + \gamma\mu u_2 - u_2^{1+1/\alpha_v}. \end{aligned}$$

Cette fonction est continue en  $u_1$  et  $u_2$ . Pour  $\gamma$  fixé, nous pouvons la décomposer en une somme de deux fonctions, une qui dépend seulement de  $u_1$  et l'autre de  $u_2$  par :

$$R_{\text{GPS}}^{(\gamma)}(u_1, u_2) = f(u_1) + g(u_2),$$

avec :

$$\begin{aligned} f(u_1) &= (1-\gamma)\mu u_1 - u_1^{1+1/\alpha_v}, \\ g(u_2) &= \gamma\mu u_2 - u_2^{1+1/\alpha_v}. \end{aligned}$$

Nous avons :

$$\begin{aligned} f'(u_1) &= (1-\gamma)\mu - (1+1/\alpha_v)u_1^{1/\alpha_v}, \\ g'(u_2) &= \gamma\mu - (1+1/\alpha_v)u_2^{1/\alpha_v}. \end{aligned}$$

Sur l'intervalle  $[1, ((1-\gamma)\mu)^{\alpha_v}]$ ,  $f$  atteint son maximum en  $\max\left(1, \left(\frac{(1-\gamma)\mu}{1+1/\alpha_v}\right)^{\alpha_v}\right)$  et, sur  $[1, (\gamma\mu)^{\alpha_v}]$ ,  $g$  atteint son maximum en  $\max\left(1, \left(\frac{\gamma\mu}{1+1/\alpha_v}\right)^{\alpha_v}\right)$ . Ainsi, comme la fonction

$R_{\text{GPS}}^{(\gamma)}(u_1, u_2)$  s'exprime par la somme de  $f$  et  $g$ , les prix qui maximisent le revenu sur le sous-espace  $V_O$  sont :

$$u_1^* = \max \left( 1, \left( \frac{(1-\gamma)\mu}{1+1/\alpha_v} \right)^{\alpha_v} \right) \quad \text{et} \quad u_2^* = \max \left( 1, \left( \frac{\gamma\mu}{1+1/\alpha_v} \right)^{\alpha_v} \right).$$

■

On étudie de la même façon sur les autres sous-domaines et particulièrement sur le sous-domaine  $D_O$  comprenant seulement du trafic de type «données».

**Lemme 3** *Les prix optimaux sur  $D_O$  sont :*

$$u_1^* = \min \left( 1, \left( \frac{(1-\gamma)\mu}{1+1/\alpha_d} \right)^{\alpha_d} \right) \quad \text{et} \quad u_2^* = \min \left( 1, \left( \frac{\gamma\mu}{1+1/\alpha_d} \right)^{\alpha_d} \right).$$

**Preuve** Le revenu du fournisseur s'exprime de la manière suivante :

$$R_{\text{GPS}}^{(\gamma)}(u_1, u_2) = (1-\gamma)\mu u_1 - u_1^{1+1/\alpha_d} + \gamma\mu u_2 - u_2^{1+1/\alpha_d}.$$

En utilisant les mêmes arguments que pour la proposition précédente, nous obtenons le résultat.

■

Notons que le sous-domaine  $M_O^{(1)}$  est non-vide si et seulement si :

$$1 \leq ((1-\gamma)\mu)^{\alpha_v} \Leftrightarrow \gamma \leq 1 - \frac{1}{\mu}.$$

De la même façon, le sous-domaine  $M_O^{(2)}$  est non-vide si et seulement si :

$$1 \leq (\gamma\mu)^{\alpha_v} \Leftrightarrow \gamma \geq \frac{1}{\mu}.$$

Ces résultats sont immédiats à partir de la figure 3.5.

**Lemme 4** *Les prix optimaux sur  $M_O^{(1)}$  sont :*

$$u_1^{(1)*} = \max \left( 1, \left( \frac{(1-\gamma)\mu}{1+1/\alpha_v} \right)^{\alpha_v} \right) \quad \text{et} \quad u_2^{(1)*} = \min \left( 1, \left( \frac{\gamma\mu}{1+1/\alpha_d} \right)^{\alpha_d} \right).$$

*Les prix optimaux sur  $M_O^{(2)}$  sont :*

$$u_1^{(2)*} = \min \left( 1, \left( \frac{(1-\gamma)\mu}{1+1/\alpha_d} \right)^{\alpha_d} \right) \quad \text{et} \quad u_2^{(2)*} = \max \left( 1, \left( \frac{\gamma\mu}{1+1/\alpha_v} \right)^{\alpha_v} \right).$$

**Preuve** Sur le sous-domaine  $M_O^{(1)}$ , le revenu du fournisseur s'écrit :

$$R_{\text{GPS}}^{(\gamma)}(u_1, u_2) = (1 - \gamma)\mu u_1 - u_1^{1+1/\alpha_v} + \gamma\mu u_2 - u_2^{1+1/\alpha_d}.$$

Sur le sous-domaine  $M_O^{(2)}$ , le revenu du fournisseur s'écrit :

$$R_{\text{GPS}}^{(\gamma)}(u_1, u_2) = (1 - \gamma)\mu u_1 - u_1^{1+1/\alpha_d} + \gamma\mu u_2 - u_2^{1+1/\alpha_v}.$$

De la même manière que pour les deux propositions précédentes, nous décomposons le revenu suivant deux fonctions dépendantes chacune d'un des prix et nous obtenons le résultat. ■

Nous sommes donc capable de déterminer, sur chacun des sous-domaines, les prix qui optimisent le revenu du fournisseur en considérant le paramètre  $\gamma$  de partage de ressource fixé. Comme pour le cas avec les classes ouvertes, nous déterminons maintenant le partage de ressource donnant le revenu maximal en considérant les prix optimaux. Nous obtenons que lorsque le système détermine ses paramètres de façon à optimiser son revenu alors, seule une unique classe de service est traitée (le paramètre  $\gamma \in \{0, 1\}$ ). Alors le système correspond simplement à une file d'attente M/M/1 avec la discipline de service PAPS.

**Proposition 4** *Le revenu maximal  $R_{\text{GPS}}^*$  du modèle approché de GPS avec classes ouvertes est égal au revenu maximal avec une politique de service PAPS.*

**preuve** Nous observons le revenu du fournisseur sur chacun des sous-domaines  $V_O$ ,  $D_O$ ,  $M_O^{(1)}$  et  $M_O^{(2)}$ . On s'intéresse tout d'abord au sous-domaine  $V_O$  pour lequel seul le trafic de type «voix» est présent dans les deux classes de service. Les prix optimaux sur ce sous-domaine sont donnés par le lemme 2 :

$$u_1^* = \max\left(1, \left(\frac{(1-\gamma)\mu}{1+1/\alpha_v}\right)^{\alpha_v}\right) \text{ et } u_2^* = \max\left(1, \left(\frac{\gamma\mu}{1+1/\alpha_v}\right)^{\alpha_v}\right).$$

Ainsi, nous avons les 4 cas suivants à étudier.

1. Supposons que  $u_1^* = \left(\frac{(1-\gamma)\mu}{1+1/\alpha_v}\right)^{\alpha_v}$  et  $u_2^* = \left(\frac{\gamma\mu}{1+1/\alpha_v}\right)^{\alpha_v}$ . A partir de l'équation (3.9), le revenu optimal en fonction du partage  $\gamma$  de ressources est :

$$R_{\text{GPS}}^*(\gamma) = (1 - \gamma)^{1+1/\alpha_v} A(\alpha_v) + \gamma^{1+1/\alpha_v} A(\alpha_v). \quad (3.9)$$

Cette fonction est strictement convexe en  $\gamma$ , et son maximum est atteint aux bords de son intervalle de définition donné par l'encadrement (3.8). Donc, le revenu maximal du fournisseur est :

$$\begin{aligned} R_{\text{GPS}}^* &= \max\left(R_{\text{GPS}}^*\left(\frac{1}{\mu}\right), R_{\text{GPS}}^*\left(1 - \frac{1}{\mu}\right)\right), \\ &\leq \max\left(R_{\text{GPS}}^*(0), R_{\text{GPS}}^*(1)\right) \leq R_{\text{PAPS}}^*. \end{aligned}$$

En effet, la fonction décrite dans l'équation (3.9) définie sur l'intervalle  $[0, 1]$  est strictement convexe et donc atteint son maximum en 0 ou en 1.

2. Supposons maintenant que  $u_1^* = 1$  et  $u_2^* = 1$ . Le revenu du fournisseur en fonction de  $\gamma$  est

$$R_{\text{GPS}}^*(\gamma) = \mu - 2.$$

De plus, pour une file avec un serveur PAPS et avec  $u = 1$ , le revenu est égal à  $\mu - 1$  (voir [73]). Ainsi, nous avons également que  $R_{\text{GPS}}^* \leq R_{\text{PAPS}}^*$ .

3. Supposons maintenant que  $u_1^* = \left(\frac{(1-\gamma)\mu}{1+1/\alpha_v}\right)^{\alpha_v}$  et  $u_2^* = 1$ . Le revenu optimal en fonction du paramètre  $\gamma$  s'écrit :

$$R_{\text{GPS}}^*(\gamma) = (1 - \gamma)^{1+1/\alpha_v} A(\alpha_v) + \gamma\mu - 1.$$

Cette fonction est également strictement convexe en  $\gamma$ . Donc, nous en déduisons que

$$R_{\text{GPS}}^* \leq \max(R_{\text{GPS}}^*(0), R_{\text{GPS}}^*(1)) \leq \max(A(\alpha_v) - 1, \mu - 1) \leq R_{\text{PAPS}}^*.$$

4. Finalement, le cas avec  $u_1^* = 1$  et  $u_2^* = \left(\frac{\gamma\mu}{1+1/\alpha_v}\right)^{\alpha_v}$  est similaire au précédent.

Une approche identique est appliquée aux sous-domaines  $D_O$ ,  $M_O^{(1)}$  et  $M_O^{(2)}$ . Sur  $D_O$ , seul le trafic de type «données» est présent dans les deux classes de service et l'analyse précédente s'applique directement en échangeant les rôles de  $\alpha_v$  et  $\alpha_d$ , avec les prix optimaux du lemme 3. Par des arguments similaires de convexité, nous obtenons également que  $R_{\text{GPS}}^* \leq R_{\text{PAPS}}^*$ .

Finalement, si l'on considère  $\gamma = 0$  ou  $\gamma = 1$  avec la politique de service GPS alors le système réagit comme une file avec politique de service PAPS car toute la capacité du serveur est allouée à une unique classe de service, laquelle admet une politique de service «premier arrivé, premier servi». Ainsi, nous déduisons que  $R_{\text{GPS}}^* = R_{\text{PAPS}}^*$ . ■

Nous avons donc déterminé les revenus optimaux dans les cas de classes dédiées et ouvertes pour le modèle avec la politique de service GPS approché par un système avec deux files d'attente indépendantes. Nous avons obtenu, pour les deux politiques de choix de la classe de service, que le revenu maximal correspond au revenu maximal avec une unique file d'attente PAPS. Grâce à cet argument, nous pouvons comparer ces revenus optimaux avec ceux obtenus avec la politique de priorité stricte de [73].

### 3.3.2 Comparaison des revenus optimaux avec GPS et avec priorités strictes

Dans cette partie, nous comparons les revenus optimaux obtenus en utilisant les deux types d'algorithmes de gestion des classes, les priorités strictes noté PQ et GPS approché par notre modèle de files indépendantes. Pour réaliser cette comparaison, nous confrontons le revenu maximal avec celui obtenu pour le modèle d'une file M/M/1 PAPS. On s'intéresse tout d'abord au modèle où chaque type de trafic est dédié à une classe de service.

### 3.3.2.1 Classes dédiées

Nous reprenons le cas de figure où le système décide de traiter le trafic de type «voix» dans la première classe de service, et le trafic de type «données» dans la seconde. Pour ce type de différenciation, nous obtenons que le revenu maximal obtenu dans le théorème 5 avec la politique de service GPS approchée est égale au revenu maximal d'une file PAPS et donc, d'après [73], est inférieur au revenu maximal obtenu avec la politique de service à priorités strictes.

**Théorème 6** *Les revenus maximaux suivant les politiques de service dans le cas des classes dédiées vérifient :*

$$R_{GPS}^* = R_{PAPS}^* \leq R_D^*,$$

où  $R_D^*$  est le revenu maximal obtenu avec un ordonnancement à priorités strictes.

En d'autres termes, le revenu maximal dans un système utilisant la politique de service à priorités strictes est toujours supérieur ou égal au revenu maximal obtenu avec la politique GPS approchée sous l'hypothèse de trafic important. De plus, le revenu maximal d'une file d'attente PAPS que pour un système système avec la politique de service GPS approchée par plusieurs files d'attente indépendantes.

**Preuve** La preuve de ce théorème se fait en deux étapes. Tout d'abord, nous montrons la relation entre les revenus maximaux avec la politique de service GPS approchée et la discipline PAPS. Ensuite, nous utilisons le résultat obtenu par Mandjes dans [73] qui compare le revenu maximal avec la politique de priorités strictes et la politique PAPS. Le revenu maximal avec la politique de service GPS approchée par deux files d'attentes indépendantes, est d'après le théorème 5 :

$$R_{GPS}^* = \max(A(\alpha_v), A(\alpha_d)),$$

avec  $A(x) = \mu^{1+x} \left(\frac{x}{1+x}\right)^x \frac{1}{1+x}$ . Ceci est équivalent à  $R_{GPS}(0)$  ou  $R_{GPS}(1)$  qui correspond au revenu optimal pour une file M/M/1 avec la politique de service PAPS et donc,

$$R_{GPS}^* = R_{PAPS}^*.$$

D'autre part, une file PAPS peut être vu comme un cas particulier d'un système avec priorités strictes ; il suffit pour cela de considérer les prix  $u_1 = \mu^{\alpha_v}$  et  $u_2 = u_1^{2\alpha_d/\alpha_v} / \mu^{\alpha_d}$ . Alors, un système utilisant une politique de service à priorités strictes ne peut engendrer un revenu plus faible qu'une file PAPS, i.e.

$$R_D^* \geq R_{PAPS}^*.$$

On en déduit les relations :

$$R_{GPS}^* = R_{PAPS}^* \leq R_D^*.$$

■

Ce théorème donne la relation entre les revenus optimaux avec la politique de service GPS approchée par deux files d'attente indépendantes et une file avec la politique de service avec priorités strictes. Nous avons donc :

$$R_{GPS}^* \leq R_D^*.$$

Ainsi, du point de vue du fournisseur de service, l'optimisation de son profit où le choix de la classe de service est décidé par le fournisseur, implique l'utilisation de la politique de gestion des classes de service avec priorités strictes par rapport à GPS.

On s'intéresse maintenant au cas où le choix de la classe de service est laissé aux utilisateurs.

### 3.3.2.2 Classes ouvertes

Comme dans le cas précédent, nous utilisons la file PAPS afin de comparer les revenus optimaux des systèmes avec politique de service à priorités strictes et GPS approchée.

**Lemme 5 (Mandjes [73])** *Dans un système avec classes ouvertes, nous avons :*

$$R_{PAPS}^* \leq R_O^*,$$

où  $R_O^*$  est le revenu maximal obtenu avec un ordonnancement avec priorités strictes.

**Idée de preuve** Il suffit de montrer, comme dans le cas avec les classes dédiées, que le modèle avec une file d'attente PAPS est obtenu en considérant des prix particuliers pour le modèle avec priorités strictes. Ainsi, le revenu maximal avec priorités strictes est supérieur ou égal au revenu maximal avec une file d'attente PAPS. ■

Nous déduisons de ces deux résultats la comparaison avec classes ouvertes, entre le revenu maximal avec la politique de service GPS approchée et le revenu maximal avec priorités strictes.

**Théorème 7** *Les revenus maximaux  $R_{GPS}^*$  et  $R_O^*$  avec classes ouvertes obtenus respectivement pour les politiques de service GPS approchée et priorités strictes vérifient :*

$$R_{GPS}^* \leq R_O^*.$$

**Preuve** Ce résultat est direct à partir de la proposition 4 et du lemme 5. ■

Ainsi, le revenu maximal obtenu dans un système avec priorités strictes est toujours supérieur ou égal au revenu obtenu dans un système avec un algorithme d'ordonnement de type GPS approché. De plus, le revenu maximal obtenu dans un modèle sans différenciation de service donne le même revenu optimal que dans un modèle avec GPS approchée.



Nous avons donc montré l'impact sur le revenu du fournisseur des différentes politiques d'ordonnancement entre classes de service. Nous nous sommes principalement intéressés à la comparaison du revenu maximal obtenu avec les politiques de service à priorités strictes et GPS. Ce dernier mécanisme a été approché par système de multiple files d'attentes indépendantes. Cette approximation est justifiée dans un contexte de trafic important où l'on suppose que la probabilité de trouver une file vide est très faible. Pour ce modèle, le revenu du fournisseur est maximum en considérant la politique de service à priorités strictes plutôt que GPS approchée.

### 3.4 Politique de service à processeur partagé de façon discriminante (DPS)

Nous nous intéressons à une autre politique de différenciation qui est le processeur partagé de façon discriminatoire ou DPS pour *Discriminatory Processor Sharing*. Cette politique de service a été introduite par Kleinrock dans [62] sous le nom de processeur partagé à priorité. La capacité du serveur attribuée à une file est partagée entre tous les paquets présents dans cette file. Ainsi, il n'y a pas de temps d'attente dans un système avec politique de service DPS, chaque paquet qui rejoint le système entre en service dès son arrivée. Ceci revient à considérer une politique PS (*Processor Sharing*) à l'intérieur de chaque file (voir figure 3.6). Le paramètre  $\gamma_i$  est ici le paramètre de priorité de la  $i^{eme}$  classe et non la proportion de ressource associée au trafic de type  $i$  comme dans GPS. Soit  $n_i$  le nombre de paquets de type  $i$  présent dans la file. Un paquet de type  $i$  est servi avec une proportion

$$\frac{\gamma_i}{\sum_{j=1}^K \gamma_j n_j}$$

du taux de service de la file. Reprenons notre modèle avec deux types de trafic en considérant les poids  $\gamma_1 = 1 - \gamma$  et  $\gamma_2 = \gamma$ . Supposons de plus qu'il y ait  $n_v$  (resp.  $n_d$ ) paquets de type «voix» (resp. «données») dans la première (resp. seconde) file. On considère seulement le cas avec classes dédiées. Alors, un paquet de type «voix» obtient un taux de service

$$\frac{(1 - \gamma)\mu}{(1 - \gamma)n_v + \gamma n_d}, \quad (3.10)$$

et un paquet de type «données»

$$\frac{\gamma\mu}{(1 - \gamma)n_v + \gamma n_d}. \quad (3.11)$$

*Remarque* La politique de service DPS n'assure pas, contrairement à GPS, un débit minimum. Seulement, le serveur traite à la suite, une partie de chaque paquet présent dans la file. Ce mécanisme est donc intéressant pour des applications élastiques qui n'ont pas de fortes contraintes de délai et particulièrement, les applications utilisant le

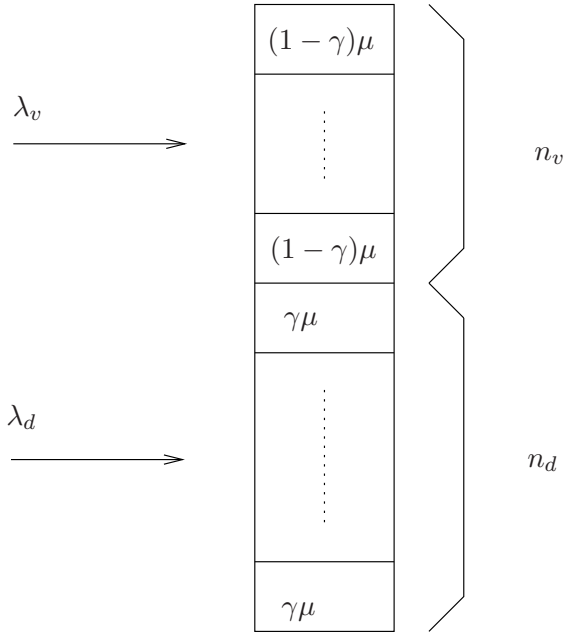


FIG. 3.6 – Modèle DPS avec  $n_d$  et  $n_v$  paquets dans le système.

protocole TCP qui adaptent le débit d'envoi des paquets en fonction de la congestion de la file.

Le délai moyen pour chaque classe de service est égal au temps moyen de service car il n'y a pas d'attente. Il existe une forme close du délai moyen pour chaque classe de service. Dans [43], les auteurs ont montré que les délais moyens pour chaque classe de service sont :

$$\mathbb{E}D_1 = \frac{1}{\mu - \lambda_1 N_1 - \lambda_2 N_2} \left( 1 + \frac{\lambda_2 N_2 (2\gamma - 1)}{\mu - (1 - \gamma)\lambda_1 N_1 - \gamma\lambda_2 N_2} \right), \quad (3.12)$$

et

$$\mathbb{E}D_2 = \frac{1}{\mu - \lambda_1 N_1 - \lambda_2 N_2} \left( 1 - \frac{\lambda_1 N_1 (2\gamma - 1)}{\mu - (1 - \gamma)\lambda_1 N_1 - \gamma\lambda_2 N_2} \right). \quad (3.13)$$

On remarque à partir de ces expressions que les nombres d'utilisateurs  $N_1$  et  $N_2$  influent sur le délai moyen de chaque classe.

### 3.4.1 Jeu non-coopératif entre classes de service

Lorsque le nombre d'utilisateurs de type  $i$  ( $i \in \{v, d\}$ ) augmente, le taux de service d'un paquet de classe  $i$  diminue. Ainsi pour cette modélisation, les fonctions d'utilité dépendent du nombre d'utilisateurs de chacune des classes de trafic et non pas de sa propre classe comme pour le modèle précédent avec l'approximation de GPS. On rappelle que chaque utilisateur décide de rejoindre ou quitter le système suivant si son

utilité est positive ou non. Cette décision implique un jeu non-coopératif entre les classes de service sur le nombre d'utilisateurs. Nous analysons ce jeu entre classes de service et montrons l'existence d'un unique équilibre de Nash qui dépend des paramètres du modèle : les prix et le paramètre de priorité. L'équilibre de Nash pour ce jeu correspond à une situation où le nombre d'utilisateurs de chaque classe est stable ; l'augmentation de ce nombre pour une des classes induit une utilité négative pour les deux. Nous avons donc un jeu dont la décision se fait au niveau de l'utilisateur dont on étudie l'équilibre au niveau des classes de service. Ce type de jeu à deux niveaux basé sur la notion d'équilibre de Stackelberg a été proposé dans [13] où chaque utilisateur choisit son débit en fonction des autres ; ce qui conduit à un équilibre de Nash, puis le système optimise son revenu avec l'équilibre sous-jacent. Dans notre étude, la dynamique est différent puisque les utilisateurs joignent ou quittent le système suivant leur utilité. L'équilibre de Nash  $(N_v^*, N_d^*)$  doit satisfaire un des points suivants :

1.  $N_v^*, N_d^* > 0$  et  $U_v(N_v^*, N_d^*) = U_d(N_v^*, N_d^*) = 0$ , aucun type d'utilisateurs n'a intérêt à rejoindre ou quitter le système,
2.  $N_v^* > 0, N_d^* = 0$ ,  $U_v(N_v^*, 0) = 0$  et  $U_d(N_v^*, 0) < 0$ , aucun utilisateur de type «voix» n'a intérêt à rejoindre ou quitter le système et aucun utilisateur de type «données» a intérêt à rejoindre car son utilité résiduelle est négative,
3.  $N_v^* = 0, N_d^* > 0$ ,  $U_d(0, N_d^*) = 0$  et  $U_v(0, N_d^*) < 0$ , correspondant à l'équilibre opposé du précédent.

### 3.4.2 Étude de l'équilibre du système

Nous nous intéressons tout d'abord à savoir s'il existe un équilibre de Nash pour ce jeu. Pour  $\gamma$  fixé, nous définissons les courbes de niveau suivantes :

$$U_v^0 = \{(N_v, N_d)/U_v(N_v, N_d) = 0\} \quad \text{et} \quad U_d^0 = \{(N_v, N_d)/U_d(N_v, N_d) = 0\}.$$

De plus, dans un souci de simplification des expressions, nous définissons les fonctions de prix  $q_v = \alpha_v \sqrt{u_v}$  et  $q_d = \alpha_d \sqrt{u_d}$ . Nous avons un premier résultat donnant la position des deux courbes de niveau  $U_v^0$  et  $U_d^0$  en fonction de  $N_v$  et  $N_d$ .

**Lemme 6** *Sur l'espace  $N_v, N_d \geq 0$ , les courbes de niveau  $U_v^0$  et  $U_d^0$  sont soit l'une au-dessus de l'autre, soit elles se croisent en un unique point  $(N_v^*, N_d^*)$  défini par :*

$$N_v^* = \left( -\frac{\mu q_v}{\gamma q_d - (1 - \gamma) q_v} + \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right) \frac{1}{\lambda_v}$$

et

$$N_d^* = \left( \frac{\mu q_d}{\gamma q_d - (1 - \gamma) q_v} - \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right) \frac{1}{\lambda_d}.$$

**Preuve** Nous considérons le système d'équations non-linéaire défini par les équations  $U_v(x, y) = 0$  et  $U_d(x, y) = 0$  et le résolvons sur  $\mathbb{R}^2$ . Ce système s'écrit :

$$(S) : \begin{cases} \frac{\mu - x\lambda_v - y\lambda_d}{q_v} = 1 + \frac{y\lambda_d(2\gamma - 1)}{\mu - (1 - \gamma)x\lambda_v - y\gamma\lambda_d} \\ \frac{\mu - x\lambda_v - y\lambda_d}{q_d} = 1 - \frac{x\lambda_v(2\gamma - 1)}{\mu - (1 - \gamma)x\lambda_v - y\gamma\lambda_d} \end{cases}$$

Il possède une unique solution  $(x^*, y^*)$  définie par :

$$x^* = \left( -\frac{\mu q_v}{\gamma q_d - (1 - \gamma)q_v} + \frac{q_v q_d}{\gamma q_v - (1 - \gamma)q_d} \right) \frac{1}{\lambda_v},$$

$$y^* = \left( \frac{\mu q_d}{\gamma q_d - (1 - \gamma)q_v} - \frac{q_v q_d}{\gamma q_v - (1 - \gamma)q_d} \right) \frac{1}{\lambda_d}.$$

L'équation  $U_d(x, y) = 0$ , tout comme l'équation  $U_v(x, y) = 0$ , est équivalente à l'écriture  $y = f(x)$  avec  $f$  une fonction continue. Nous avons donc deux courbes de fonctions continues qui se croisent en un unique point  $(x^*, y^*)$  de  $\mathbb{R}^2$ . Donc, si  $x^*$  est négatif ou  $y^*$  est négatif, les courbes sont l'une au-dessus de l'autre sur  $\mathbb{R}_+^2$ . Sinon, elles se croisent en  $(x^*, y^*) \in \mathbb{R}_+^2$ . ■

D'après ce lemme, nous avons trois configurations possibles des positions des courbes de niveau  $U_v^0$  et  $U_d^0$  l'une par rapport à l'autre. Nous montrons dans la suite que chaque configuration induit un équilibre de Nash différent et unique. Nous décrivons dans chaque configuration des courbes, la dynamique du système où les individus partent ou arrivent suivant leur utilité et qui amène à l'équilibre. Nous supposons  $U_v(0, 0) > 0$  et  $U_d(0, 0) > 0$  afin qu'au moins un utilisateur de n'importe quel type de trafic possède un intérêt à rejoindre le système s'il est vide. Le théorème suivant détermine l'équilibre de Nash  $(N_v^*, N_d^*)$  du système qui stabilise le nombre d'utilisateurs de chaque type en fonction des paramètres qui sont : le taux de service  $\mu$ , le paramètre de priorité  $\gamma$  et les prix  $q_v$  et  $q_d$ . Ces derniers sont supposés vérifier  $q_v < \mu$  et  $q_d < \mu$  afin que chaque type de trafic ait un intérêt à rejoindre le système. En effet, supposons que  $q_i \geq \mu$  avec  $i \in \{v, d\}$ .

- Si  $N_j = 0$  avec  $j \in \{v, d\}$  et  $j \neq i$ , alors pour tout  $N_i > 0$ ,

$$\begin{aligned} U_i(\mathbb{E}D_i(N_i, 0)) &= \frac{1}{\mathbb{E}D_i(N_i, 0)^{\alpha_i}} - u_i \\ &= (\mu - \lambda_i N_i)^{\alpha_i} - u_i < 0, \end{aligned}$$

car  $\mu - q_i < 0 < \lambda_i N_i$ . Alors le nombre d'utilisateurs de type  $i$  est nul car l'utilité résiduelle est négative pour tout  $N_i > 0$ .

- Sinon,  $N_j > 0$  alors  $\mathbb{E}D_i(N_i, N_j) > \mathbb{E}D_i(N_i, 0)$ . Or l'utilité résiduelle est décroissante et donc :

$$U_i(\mathbb{E}D_i(N_i, N_j)) < U_i(\mathbb{E}D_i(N_i, 0)) < 0,$$

d'après le cas précédent.

**Théorème 8** Nous supposons que le taux de service global  $\mu$  et le paramètre de priorité  $\gamma$  sont fixés. Nous posons  $q_d^- = \sqrt{\mu q_v + \left(\frac{\mu - q_v}{2} \frac{1 - \gamma}{\gamma}\right)^2} - \frac{\mu - q_v}{2} \frac{1 - \gamma}{\gamma}$ ,  $q_d^+ = \frac{\mu \gamma q_v + (1 - \gamma) q_v^2}{\gamma q_v + \mu(1 - \gamma)}$  alors pour tout  $q_v < \mu$  :

1. Si  $0 < q_d < q_d^-$  alors l'équilibre est  $(N_v^*, N_d^*) = \left(\frac{\mu - q_v}{\lambda_v}, 0\right)$ ,
2. si  $q_d^- < q_d < q_d^+$  alors l'équilibre est  $(N_v^*, N_d^*)$ ,
3. si  $q_d^+ < q_d$  alors l'équilibre est  $(N_v^*, N_d^*) = \left(0, \frac{\mu - q_d}{\lambda_d}\right)$ ,

avec

$$N_d^* = \frac{1}{\lambda_d} \left( \frac{\mu q_d}{\gamma q_d - (1 - \gamma) q_v} - \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right), \quad (3.14)$$

et

$$N_v^* = \frac{1}{\lambda_v} \left( -\frac{\mu q_v}{\gamma q_d - (1 - \gamma) q_v} + \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right). \quad (3.15)$$

Afin de montrer ce théorème, nous avons besoin du lemme suivant.

**Lemme 7** Nous gardons les notations du théorème précédent et de plus nous considérons la solution  $(x^*, y^*)$  donnée dans la preuve du lemme 6. Alors :

$$x^* > 0 \Leftrightarrow q_d < q_d^+, \quad \text{et} \quad y^* > 0 \Leftrightarrow q_d > q_d^-.$$

**Preuve** La solution  $(x^*, y^*)$  du système (S) est donnée dans le lemme 6 par :

$$x^* = \left( -\frac{\mu q_v}{\gamma q_d - (1 - \gamma) q_v} + \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right) \frac{1}{\lambda_v},$$

$$y^* = \left( \frac{\mu q_d}{\gamma q_d - (1 - \gamma) q_v} - \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right) \frac{1}{\lambda_d}.$$

Nous avons :

$$x^* > 0 \Leftrightarrow \frac{\mu}{\gamma q_v - (1 - \gamma) q_d} < \frac{q_d}{\gamma q_v - (1 - \gamma) q_d},$$

$$\Leftrightarrow \gamma q_d - (1 - \gamma) q_v < \mu \gamma - \mu(1 - \gamma) \frac{q_d}{q_v},$$

$$\Leftrightarrow q_d < \frac{(1 - \gamma) q_v^2 + \mu \gamma q_v}{\gamma q_v + \mu(1 - \gamma)} = q_d^+.$$

De même, nous avons :

$$y^* > 0 \Leftrightarrow \frac{q_v}{\gamma q_v - (1 - \gamma) q_d} < \frac{\mu}{\gamma q_v - (1 - \gamma) q_d},$$

$$\Leftrightarrow \mu \gamma \frac{q_v}{q_d} - (1 - \gamma) \mu < \gamma q_d - (1 - \gamma) q_v,$$

$$\Leftrightarrow q_d > \sqrt{\mu q_v + \left(\frac{\mu - q_v}{2} \frac{1 - \gamma}{\gamma}\right)^2} - \frac{\mu - q_v}{2} \frac{1 - \gamma}{\gamma} = q_d^-.$$

■

Nous pouvons maintenant prouver le théorème 8.

**Preuve du théorème 8** Définissons tout d'abord pour tout  $q_v < \mu$  :

$$q_d^- = \sqrt{\mu q_v + \left(\frac{\mu - q_v}{2} \frac{1 - \gamma}{\gamma}\right)^2} - \frac{\mu - q_v}{2} \frac{1 - \gamma}{\gamma} \quad \text{et} \quad q_d^+ = \frac{\mu \gamma q_v + (1 - \gamma) q_v^2}{\gamma q_v + \mu(1 - \gamma)}.$$

- Supposons tout d'abord que  $0 < q_d < q_d^-$ , alors la solution  $(x^*, y^*)$  du système  $(S)$  vérifie  $x^* > 0$  et  $y^* < 0$  d'après le lemme 7. Ainsi, la configuration des courbes de niveau est représentée sur la figure 3.7 (a) où  $U_d^0$  est en-dessous  $U_v^0$ . Comme  $U_v(0, 0) > 0$  et  $U_d(0, 0) > 0$ , le nombre d'utilisateur de type «voix»  $N_v$  et «données»  $N_d$  augmente jusqu'à atteindre le niveau  $U_d^0$ . Pour ce point, l'application «données» n'a plus intérêt à changer son nombre d'utilisateurs mais  $U_v(N_v, N_d) > 0$  et donc  $N_v$  continue à augmenter. Ainsi, nous passons dans la zone entre les deux courbes de niveau avec  $U_d(N_v, N_d) < 0$  et  $U_v(N_v, N_d) > 0$ . Dès lors, le nombre de sources de type «données» diminue afin d'annuler son utilité et le nombre de source de type «voix» augmente. Cette dynamique entre le nombre d'utilisateurs de chaque type qui consiste à glisser en augmentant  $N_v$  et diminuant  $N_d$  le long de la courbe de niveau  $U_d^0$  et conduit à  $N_d = 0$ . Finalement, le nombre d'utilisateurs de type «voix» augmente afin d'annuler l'utilité et le jeu atteint son équilibre :

$$(N_v^*, N_d^*) = \left( \frac{\mu - q_v}{\lambda_v}, 0 \right).$$

Le nombre d'utilisateurs de type «données» à l'équilibre du système est nul.

- On suppose maintenant que  $q_d^+ < q_d < \mu$ , alors d'après le lemme 7, la solution  $(x^*, y^*)$  du système  $(S)$  vérifie  $x^* < 0$  et  $y^* > 0$ . Donc, cette configuration est donnée par la figure 3.7 (b). La dynamique du jeu est semblable à la configuration précédente et l'équilibre du système est :

$$(N_v^*, N_d^*) = \left( 0, \frac{\mu - q_d}{\lambda_d} \right).$$

- Enfin, si  $q_d^- < q_d < q_d^+$  alors, d'après le lemme 7, la solution  $(x^*, y^*)$  du système  $(S)$  vérifie  $x^* > 0$  et  $y^* > 0$ . Ainsi, nous sommes dans la configuration représentée sur la figure 3.7 (c), où les courbes de niveaux  $U_d^0$  et  $U_v^0$  se croisent en un unique point  $(x^*, y^*)$ . Dans cette configuration nous avons nécessairement  $U_v^0$  au-dessus de  $U_d^0$  si et seulement si  $N_v < x^*$  et  $N_d > y^*$  car, après quelques manipulations nous avons  $U_v\left(0, \frac{\mu - q_d}{\lambda_d}\right) > 0$  et  $U_d\left(\frac{\mu - q_v}{\lambda_v}, 0\right) > 0$ . Le nombre d'utilisateurs (partant de  $(0, 0)$ ) de chaque type augmente jusqu'à atteindre la courbe minimum. Nous avons deux cas de figure possibles :
  - soit on atteint d'abord un point  $(N_v, N_d)$  tel que  $U_v(N_v, N_d) = 0$  (voir figure 3.8 la courbe en pointillé 1). Alors  $U_d(N_v, N_d) > 0$ , donc le nombre de sources de type «données» augmente et ainsi  $U_v(N_v, N_d)$  devient négative, ce qui fait diminuer le nombre de sources de type «voix».
  - Soit on atteint un point  $(N_v, N_d)$  tel que  $U_d(N_v, N_d) = 0$  (voir figure 3.8 la courbe en pointillé 2). De la même façon que précédemment,  $U_v(N_v, N_d)$  est

strictement positive, donc le nombre de sources de type «voix» augmente et ainsi  $U_d(N_v, N_d)$  devient négative, ce qui fait diminuer le nombre de sources de type «données».

Ainsi, le jeu atteint l'équilibre qui est l'intersection des courbes de niveau :

$$N_d^* = \left( \frac{\mu q_d}{\gamma q_d - (1 - \gamma) q_v} - \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right) \frac{1}{\lambda_d},$$

et

$$N_v^* = \left( -\frac{\mu q_v}{\gamma q_d - (1 - \gamma) q_v} + \frac{q_v q_d}{\gamma q_v - (1 - \gamma) q_d} \right) \frac{1}{\lambda_v}.$$

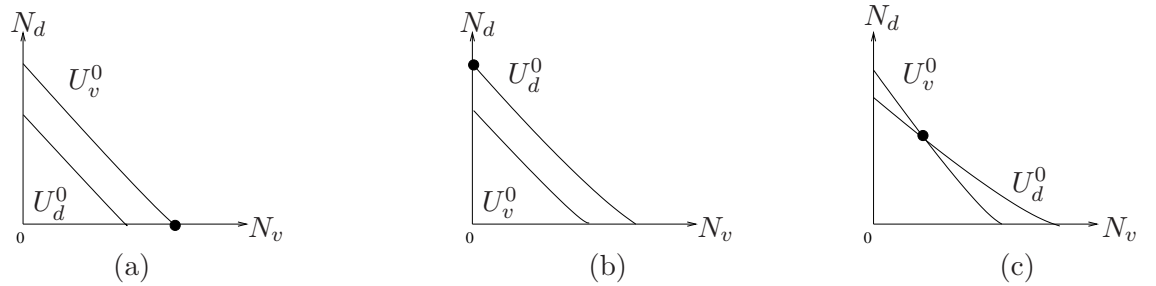


FIG. 3.7 – Les trois différentes configurations des courbes de niveau et l'équilibre de Nash associé.

*Remarque* On peut vérifier par de simples calculs que la condition de stabilité de la file d'attente, i.e.  $\rho := \frac{\lambda_v N_v^* + \lambda_d N_d^*}{\mu} < 1$  est satisfaite à l'équilibre de Nash du système.

*Exemples* La figure 6.5 montre deux exemples pour des valeurs de paramètres de priorité  $\gamma = 0.2$  et  $\gamma = 0.9$ , de la nature du trafic dans le système en fonction des prix  $q_v$  et  $q_d$ . La zone hachurée correspond au cas où les deux types de trafic, «voix» et «données», sont présents. Nous remarquons tout d'abord que lorsque le prix pour les sessions de type «voix» est très supérieur au prix pour les sessions de type «données», alors seul le trafic de type «données» sera présent dans la file (par exemple avec  $\gamma = 0.2$ , les prix  $q_v = 0.9$  et  $q_d = 0.1$ ) et inversement. Enfin, si l'on s'intéresse à l'équilibre non-trivial pour lequel les deux types de trafic sont présents à l'équilibre, si  $\gamma < 0.5$  alors les prix doivent vérifier nécessairement  $q_d < q_v$  et inversement.

Le théorème 8 détermine suivant  $q_v$ ,  $q_d$  et le paramètre de priorité  $\gamma$ , l'équilibre induit par la dynamique du jeu non-coopératif entre les types d'application. Le système s'intéresse maintenant à déterminer ces paramètres afin d'optimiser son revenu.

### 3.4.3 Optimisation du revenu du fournisseur

Le système cherche à déterminer  $u_d^*$ ,  $u_v^*$  et  $\gamma^*$  qui optimisent son revenu. Ce problème d'optimisation s'écrit sous la forme suivante :

$$(u_d^*, u_v^*, \gamma^*) = \arg \max_{u_d, u_v, \gamma} \lambda_d N_d^*(u_d, u_v, \gamma) u_d + \lambda_v N_v^*(u_d, u_v, \gamma) u_v$$

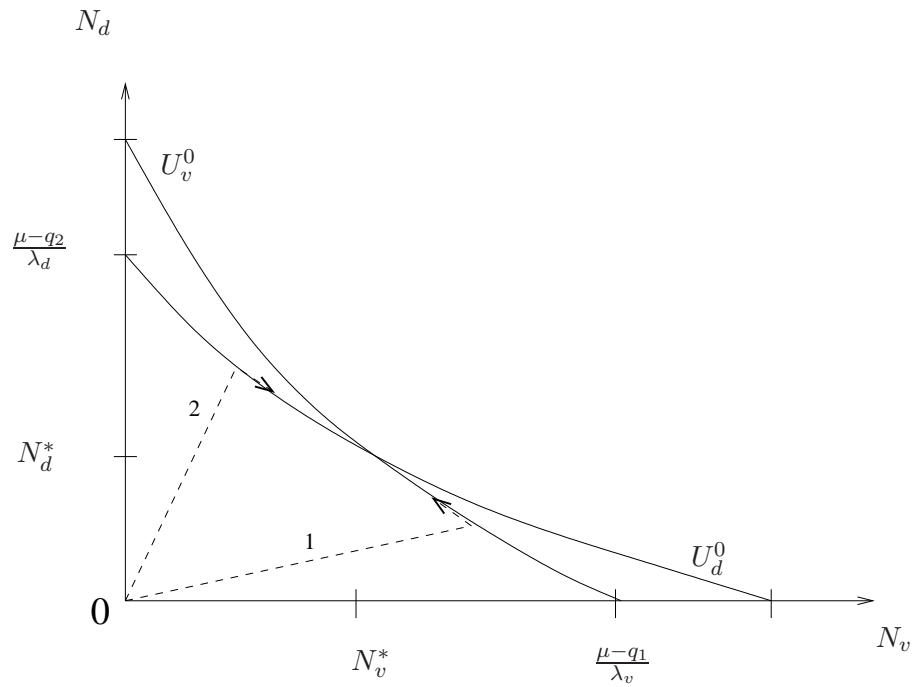


FIG. 3.8 – Dynamique du système lorsque l'équilibre de Nash non trivial est atteint.

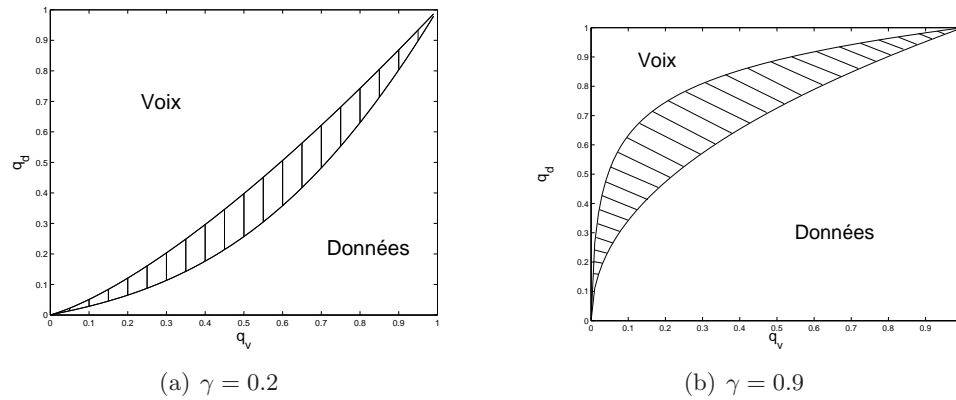


FIG. 3.9 – Types de paquets dans le système suivant les prix. Dans la zone hachurée, les deux types de paquets sont présents.



avec  $u_v, u_d \geq 0$  et  $\gamma \in [0, 1]$ .

Il est possible de déterminer de façon analytique la valeur optimale de  $\gamma$  pour des prix  $u_v$  et  $u_d$  fixés. Cependant, la détermination des prix optimaux semble irréalisable de manière théorique en raison de la complexité du système des équations du premier ordre. Nous ne décrivons pas dans ce manuscrit l'analyse théorique de l'optimisation du revenu en fonction de  $\gamma$  avec prix fixés car, une analyse numérique réalisée avec n'importe quel logiciel d'optimisation standard donne les prix et paramètres de priorité optimaux.

*Exemple* Considérons l'exemple avec les paramètres suivants :

$$\begin{aligned}\mu &= 1, \\ \alpha_v &= 1.8, \\ \alpha_d &= 1.5, \\ \lambda_v &= 0.1, \\ \lambda_d &= 0.3.\end{aligned}$$

La figure 3.10 donne le paramètre de priorité optimal  $\gamma^*$  en fonction des prix  $u_v$  et  $u_d$ . Nous remarquons que lorsque  $u_v$  est grand et  $u_d$  petit, le revenu est maximisé en prenant  $\gamma^* = 0$ , c'est-à-dire en considérant la politique de service à priorité stricte pour le trafic de type «voix» sur le trafic de type «données». En effet, si  $\gamma = 0$ , le délai moyen de chaque classe de service correspond au délai moyen pour une file avec priorités strictes de la première classe sur la seconde. Inversement, lorsque  $u_v$  est petit et  $u_d$  est important, on obtient  $\gamma^* = 1$  et donc le modèle se comporte comme une file à priorités strictes.

La figure 3.11 montre le revenu en fonction des prix  $u_v$  et  $u_d$  avec le paramètre de priorité optimal  $\gamma^*$  déterminé auparavant. Nous observons que l'enveloppe de cette courbe est régulière et qu'il est donc aisé d'obtenir le maximum global avec un algorithme d'optimisation basé sur le gradient. Nous obtenons les prix optimaux  $u_d^* = 0.24$  et  $u_v^* = 0.68$  et le paramètre de priorité optimal est  $\gamma^* = 0$ . Le revenu optimal est  $R^* = 0.21$ .

#### 3.4.4 Comparaison des revenus optimaux avec DPS et priorités strictes

Il est important de remarquer que pour l'exemple précédent, le paramètre de priorité optimal est  $\gamma^* = 0$ . Ce qui signifie que pour optimiser son revenu, le système utilise un ordonnancement à priorités strictes de la classe 1 sur la classe 2. Nous avons observé ce phénomène pour l'ensemble de nos applications numériques. Ceci implique la conjecture suivante :

**Conjecture 1** *Dans système à différenciation de service avec un algorithme d'ordonnancement à processeur partagé de façon discriminante DPS, le revenu du fournisseur est maximisé en considérant la politique de service à priorités strictes ( $\gamma^* \in \{0, 1\}$ ).*

Ceci se rapproche du résultat connue sous le nom de "c $\mu$ -rule" [64, page 125]. Cet ordonnancement détermine la meilleure façon de répartir les niveaux de priorités entre

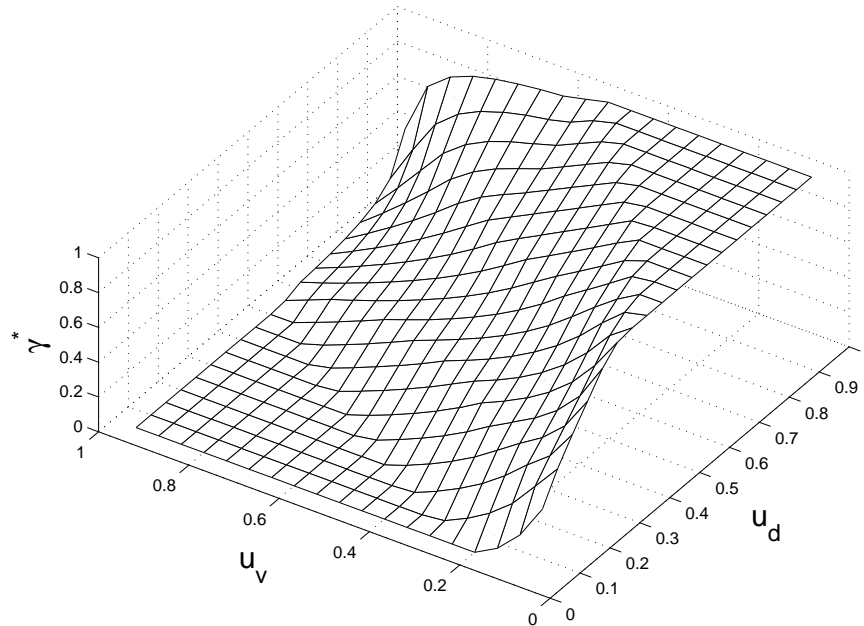


FIG. 3.10 – Paramètre de priorité optimal  $\gamma^*$  entre les classes de service en fonction des prix  $u_v$  et  $u_d$ .

les types d'utilisateurs afin de minimiser le délai moyen et de maximiser le revenu [50]. Ce mécanisme consiste à donner un ordre de priorité à la classe  $i$  décroissant suivant l'index  $c_i\mu$  où  $c_i$  est le coût marginal du délai. Ce mécanisme est prouvé être optimal dans le cas où les utilisateurs possèdent une utilité **linéaire** en fonction du délai moyen. Dans [111], il est proposé une extension de ce mécanisme, la  $Gc\mu$  rule, pour le cas des fonctions de coûts de délai convexe et d'utilité linéaire. Ce dernier mécanisme ne peut pas être appliqué à notre modèle.

### 3.4.5 Application aux trafics TCP

L'algorithme DPS possède un intérêt particulier dans la communauté des réseaux. En effet, il permet de modéliser le comportement de multiples flux élastiques hétérogènes se partageant un nœud du réseau voir [7, 26]. Le mécanisme DPS est une généralisation multiclassés de PS, lequel modélise des connections TCP homogènes (même RTT) [16, 66]. Ainsi, grâce au paramètre de priorité  $\gamma$ , nous différencions les sessions TCP. Le mécanisme d'ordonnancement DPS est considéré comme un paradigme convenable pour analyser au niveau des flux des connections TCP hétérogènes en concurrence [85]. La modélisation ne se situe plus au niveau paquets mais au niveau des sessions TCP et  $N_v$ ,  $N_d$  représentent alors le nombre moyen de sessions TCP de type «voix» et de type «données». On suppose maintenant que  $\lambda_v$  et  $\lambda_d$  représentent les tailles moyennes des sessions et qu'elles sont indépendantes. Ainsi, avec ces notations, les nombres de sessions

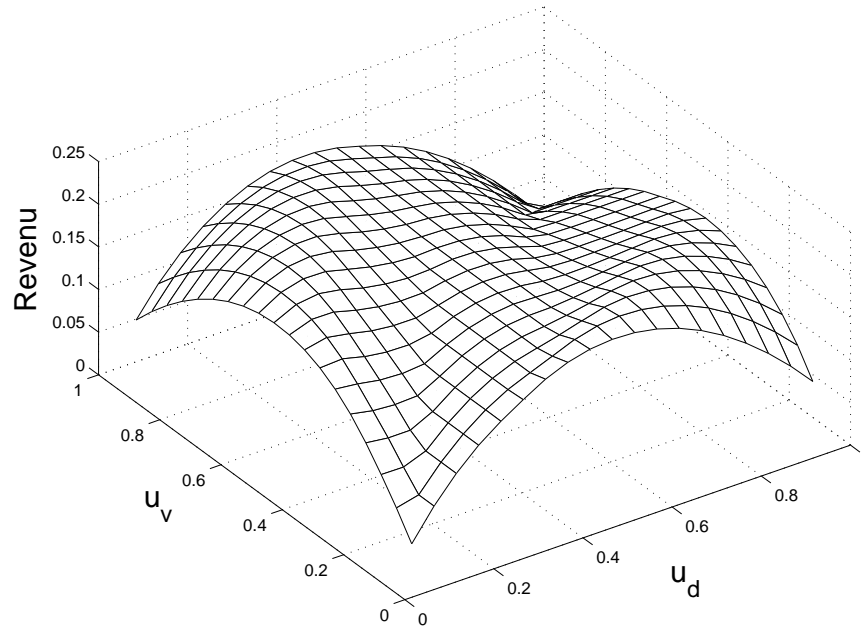


FIG. 3.11 – Revenu du fournisseur avec paramètre de priorité optimal en fonction des prix  $u_v$  et  $u_d$ .

$N_v$  de type «voix» et  $N_d$  de type «données» correspondent au nombre de clients dans une file M/M/1 avec la politique de service à processeur partagé DPS. Nous pouvons alors appliquer tous nos résultats dans le cas où  $N_v^*$  et  $N_d^*$  correspondent au nombre de sessions TCP pour des applications de type «voix» et de type «données» à l'équilibre du système. Cette étude du modèle de tarification à l'usage pour le contrôle de sessions TCP a été publiée dans [54].

### 3.5 Conclusion

Dans ce chapitre, nous avons étudié l'impact économique du mécanisme d'ordonnancement sur des systèmes multiclassés à différenciation de service. Nous avons tout d'abord considéré l'algorithme d'ordonnancement GPS qui permet de réaliser de la différenciation de service tout en garantissant un taux de service minimum à chaque classe. Ensuite, nous avons analysé un modèle utilisant l'algorithme DPS, une variation multi-classes de PS. Ce dernier mécanisme est utilisé comme modèle de multiple connexions TCP se partageant un nœud du réseau. Nous avons comparé le revenu maximal obtenu avec ces algorithmes d'ordonnancement et le mécanisme à priorités strictes. Nous avons observé que la politique à priorités strictes donne un meilleur revenu maximal au fournisseur. Ainsi, un fournisseur de service ayant la possibilité de choisir son mécanisme d'ordonnancement, aura intérêt à promouvoir de la différenciation de ser-

vices avec la politique d'ordonnancement à priorités strictes laquelle est optimale au sens du revenu par rapport à GPS et DPS.

## Chapitre 4

# Mécanisme de partage du coût de congestion et priorités strictes

Dans cette partie, nous considérons un système à différenciation de service muni du mécanisme d’ordonnancement à priorités strictes sur lequel nous nous focalisons puisque nous avons montré dans le chapitre précédent que ce mécanisme est optimal au sens du revenu du fournisseur. Ce mécanisme d’ordonnancement est largement considéré dans les modèles de tarification de l’Internet comme dans [75, 81] par exemple. De nombreux modèles de tarification issus du modèle de Mendelson et Wang [81] ont été proposés pour ce genre de problème. Nous considérons le modèle avec les classes dédiées où le système décide l’ordonnancement des classes de service. Entre tous les mécanismes à priorités, la politique d’ordonnancement à priorités strictes nonpréemptives  $c - \mu$  rule [64] minimise le coût total en fonction du délai moyen. Nous considérons dans ce chapitre cette politique à priorités strictes comme dans [81]. Les modèles de tarification basés sur cette politique d’ordonnancement utilisent une tarification au prix marginal, i.e. chaque utilisateur paye le coût perçu par les autres du fait de sa présence dans le système. Ce mécanisme est optimal au sens du surplus collectif mais il a été montré dans [84] qu’il ne partage pas le coût global perçu par le système entre les classes. Nous proposons dans ce chapitre, un mécanisme de tarification qui, en plus d’être optimal au sens du surplus collectif, partage le coût de congestion global du système entre les différentes classes de service.

Il existe en micro-économie un modèle de partage de coût appelé le mécanisme de Aumann-Shapley [108]. Ce mécanisme proposé par Aumann et Shapley fût appliqué la première fois par Billera et al. dans [18] afin de partager le coût de service téléphonique entre les utilisateurs. Il est aussi utilisé dans [51] afin d’allouer les coûts de congestion dans un système multi-classes avec priorités strictes où les taux d’arrivée sont arbitraires. Nous utilisons ici également le mécanisme de Aumann-Shapley mais, nous déterminons les taux d’arrivée optimaux au sens du surplus collectif du système. Pour cela, nous introduisons une fonction représentant la valuation du taux de service et la fonction de demande reliant ces taux d’arrivées au prix.

Nous étendons ensuite nos résultats à un réseau de Jackson ouvert et à un modèle

de tarification dynamique (la demande varie suivant les périodes de temps et le prix également suivant la demande). Ces extensions ont été proposées dans [78] pour la fonction de coût marginal.

## 4.1 Modélisation

Nous considérons un système composé de  $K$  classes de service et modélisé par une file d'attente M/M/1. Le processus d'arrivée des paquets des utilisateurs de classe  $i$  suit une loi de Poisson de taux  $\lambda_i$ , pour  $i = 1, \dots, K$ , et dépend du prix. Le temps de service  $\tau_i$  requis par un paquet de classe  $i$  suit une loi exponentielle de moyenne  $1/\mu_i$ . Ces temps de service sont supposés indépendants entre les jobs et identiquement distribués pour chaque classe. Un paquet de classe  $i$  est caractérisé par un coût  $v_i$  par unité de temps d'attente (délai) dans la file. Le modèle est basé sur une politique de service à priorités non-préemptives et les niveaux de priorités sont ordonnés suivant la  $c - \mu$  rule [64], c'est-à-dire :

$$v_1\mu_1 \geq v_2\mu_2 \geq \dots \geq v_n\mu_n, \quad (4.1)$$

donnant la plus haute priorité à la classe 1, la seconde priorité à la classe 2 et ainsi de suite. Nous avons un système avec classes dédiées où le système décide l'ordre des classes de service suivant leur coût en fonction du délai. Nous notons  $\mathbb{E}D_i(\underline{\lambda})$  le délai moyen perçu par un job de classe  $i$  avec  $\underline{\lambda} = (\lambda_1, \dots, \lambda_n)$  le vecteur des taux d'arrivée. D'après [63], le délai moyen s'exprime en fonction des taux d'arrivée et taux de service par :

$$\forall i = 1, \dots, K, \quad \mathbb{E}D_i(\underline{\lambda}) = \frac{\sum_{j=1}^K \lambda_j / \mu_j^2}{(1 - \sum_{j=1}^{i-1} \lambda_j / \mu_j)(1 - \sum_{j=1}^i \lambda_j / \mu_j)} + 1/\mu_i. \quad (4.2)$$

Définissons par  $L_i(\underline{\lambda})$  le nombre moyen de jobs de classe  $i$  dans le système ; la formule de Little donne :

$$\forall i = 1, \dots, K, \quad L_i(\underline{\lambda}) = \lambda_i \mathbb{E}D_i(\underline{\lambda}).$$

Le coût de congestion global du système  $L(\underline{\lambda})$  est défini par :

$$L(\underline{\lambda}) = \sum_{i=1}^K v_i \lambda_i \mathbb{E}D_i(\underline{\lambda}) \quad (4.3)$$

qui représente le coût total en terme de délai par unité de temps perçu par les paquets dans le système.

(Re)-Définissons une fonction  $V_i(\lambda_i)$  spécifiant la valuation agrégée des utilisateurs de classe  $i$  par unité de temps pour un taux d'arrivée  $\lambda_i$ . Pour tout  $i$ , la fonction  $V_i(\cdot)$  est supposée croissante, dérivable et strictement concave. Le nombre d'utilisateurs de chaque classe est supposé élevé, ce qui est le cas pour le réseau Internet où le nombre de personnes utilisant le même type d'applications peut être important. Ainsi, la valuation marginale (d'un utilisateur) de la  $i^{eme}$  classe de service est représentée par  $V_i'(\lambda_i)$  la

dérivée de la valuation globale de la classe. La valuation totale du système est définie comme la somme sur l'ensemble des classe de service des valuations :

$$V(\underline{\lambda}) = \sum_{i=1}^K V_i(\lambda_i).$$

Le problème du système est de déterminer une tarification qui optimise le surplus collectif du système  $V(\underline{\lambda}) - L(\underline{\lambda})$ . Ce problème d'optimisation peut se formuler par :

$$\underline{\lambda}^* = \arg \max_{\underline{\lambda}} (V(\underline{\lambda}) - L(\underline{\lambda})). \quad (4.4)$$

Les utilisateurs de classe  $i$  sont supposés sensibles à leur prix  $u_i$  et au délai moyen  $\mathbb{E}D_i$  de transmission de leurs paquets. En absence de mécanisme de tarification, le surplus marginal d'un utilisateur de classe  $i$  est

$$V'_i(\lambda_i) - v_i \mathbb{E}D_i(\underline{\lambda}).$$

Un utilisateur rentre dans le système si et seulement si sa valuation marginale excède son coût, i.e. son surplus est positif. L'objectif du système est d'introduire une tarification (un prix) afin d'optimiser le surplus collectif du système donné par l'équation (4.4). Pour cela, le système détermine le vecteur  $\underline{u}$  des prix en utilisant le mécanisme de Aumann-Shapley [108]. Ce mécanisme de partage de coût définit le prix pour une classe comme la moyenne uniforme du coût marginal du système induit par cette classe. En considérant le surplus marginal d'une classe comme le coût marginal du système induit par celle-ci, le mécanisme de Aumann-Shapley donne :

$$u_i = \int_0^1 (V'_i(\lambda_i t) - v_i \mathbb{E}D_i(\underline{\lambda} t)) dt \quad (4.5)$$

Après quelques manipulations, l'équation (4.5) permet d'exprimer les taux d'arrivées en fonction des prix. Nous avons donc la relation de demande suivante :

$$\int_0^1 V'_i(\lambda_i t) dt = u_i + v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda} t) dt. \quad (4.6)$$

L'objectif du système est de déterminer les prix afin d'assurer que les débits de chaque classe de service maximisent le surplus collectif du système. La relation de demande (4.6) permet au système de savoir comment les utilisateurs adaptent leur débit en fonction du prix. Ainsi, le système est capable de résoudre le problème d'optimisation du surplus collectif en déterminant les débits optimaux pour chaque classe de service et ensuite, en déduire les prix optimaux avec la relation de demande (4.6). Nous résolvons tout d'abord ce problème dans le cas où les temps de service sont homogènes, puis nous montrons que cette tarification force les utilisateurs à choisir correctement leur classe de service.

## 4.2 Temps de service homogènes

Nous considérons tout d'abord que les taux de service sont égaux entre toutes les classes, i.e.  $\mu_1 = \dots = \mu_n = \mu$ . De plus, dans ce cas de figure, nous supposons, sans perte de généralités que  $\mu = 1$ .

### 4.2.1 Tarification optimale

L'objectif du système est de déterminer les prix  $u_i^*$  afin que les taux d'arrivées qui résultent de la relation de demande (4.6), maximisent le surplus collectif. En ce sens, les prix sont optimaux et définis dans le théorème suivant.

**Théorème 9** *Le prix optimal  $u_i^*$  pour la classe  $i$  avec  $i \in \{1, \dots, K\}$ , est donné par*

$$\begin{aligned}
u_i^* &= \sum_{j=1}^K v_j \lambda_j^* \int_0^1 t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\underline{\lambda}^* t) dt & (4.7) \\
&= \sum_{j=1}^K v_j \lambda_j^* (-\ln(1 - S_j)) \left( \frac{1}{\lambda_j^* S_j} + \mathbb{1}_{i \leq j} \frac{S_K(\lambda_j^{*2} - S_j^2)}{S_j^2 S_{j-1} \lambda_j^{*2}} + \mathbb{1}_{i \leq j-1} \frac{S_K(\lambda_j^{*2} - S_{j-1}^2)}{S_j S_{j-1}^2 \lambda_j^{*2}} \right) \\
&\quad + \sum_{j=1}^K v_j \lambda_j^* (-\ln(1 - S_{j-1})) \left( -\frac{1}{\lambda_j^* S_{j-1}} + \mathbb{1}_{i \leq j} \frac{S_K}{\lambda_j^{*2} S_{j-1}} + \mathbb{1}_{i \leq j-1} \frac{S_K}{\lambda_j^{*2} S_j} \right) \\
&\quad + \sum_{j=1}^K v_j \lambda_j^* \left( \mathbb{1}_{i \leq j} \frac{S_K}{\lambda_j^* S_j (1 - S_j)} - \mathbb{1}_{i \leq j-1} \frac{S_K}{\lambda_j^* S_{j-1} (1 - S_{j-1})} \right).
\end{aligned}$$

avec  $S_j = \sum_{i=1}^j \lambda_i^*$  et  $\mathbb{1}_{i \leq j}$  la fonction indicatrice de l'événement  $\{i \leq j\}$ .

**Preuve** On s'intéresse à la  $i^{\text{eme}}$  classe de service. La condition du premier ordre du problème d'optimisation (4.4) par rapport à  $\lambda_i$  donne la relation pour le débit optimal  $\underline{\lambda}^*$  :

$$V_i'(\lambda_i^*) = \sum_{j=1}^K v_j \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda}^*). \quad (4.8)$$

De plus, la relation de demande (4.6) donne

$$\begin{aligned}
\int_0^1 V_i'(\lambda_i t) dt &= u_i + v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda} t) dt, \\
&= \int_0^1 \sum_{j=1}^n v_j \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda} t) dt.
\end{aligned}$$

En considérant la relation avec le débit optimal donné par l'équation (4.8), le prix optimal  $u_i^*$  s'écrit

$$u_i^* = \int_0^1 \sum_{j=1}^K v_j \frac{\partial L_j}{\partial \lambda_i}(\underline{\lambda}^* t) dt - v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda}^* t) dt.$$



D'après la formule de Little, nous avons  $L_i(\lambda) = \lambda_i \mathbb{E}D_i(\lambda)$ . De plus, nous avons

$$\frac{\partial L_i}{\partial \lambda_i}(\lambda) = \mathbb{E}D_i(\lambda) + \lambda_i \frac{\partial \mathbb{E}D_i}{\partial \lambda_i}(\lambda) \quad \text{et} \quad \frac{\partial L_j}{\partial \lambda_i}(\lambda) = \lambda_j \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\lambda) \quad \text{pour } j \neq i,$$

qui donne

$$\begin{aligned} u^* &= \int_0^1 \sum_{j \neq i} v_j \lambda_j^* t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\lambda^* t) dt + v_i \int_0^1 \left( \mathbb{E}D_i(\lambda^* t) + \lambda_i^* t \frac{\partial \mathbb{E}D_i}{\partial \lambda_i}(\lambda^* t) \right) dt \\ &\quad - v_i \int_0^1 \mathbb{E}D_i(\lambda^* t) dt, \\ &= \sum_{j=1}^K v_j \lambda_j^* \int_0^1 t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\lambda^* t) dt, \\ &= \sum_{j=1}^K v_j \lambda_j^* \int_0^1 t \left( \frac{1}{(1-tS_j)(1-tS_{j-1})} + \frac{\mathbb{1}_{i \leq j} \mathbb{E}D_j^q}{1-tS_j} + \frac{\mathbb{1}_{i \leq j-1} \mathbb{E}D_j^q}{1-tS_{j-1}} \right) dt, \end{aligned}$$

avec  $\mathbb{E}D_j^q(\lambda^* t) = \frac{t \sum_{i=1}^K \lambda_i^*}{(1-tS_j)(1-tS_{j-1})}$  le temps d'attente dans le système (temps de réponse moins le temps moyen de service). Les décompositions en éléments simples par rapport à  $t$  donnent

$$\frac{t}{(1-tS_j)(1-tS_{j-1})} = \frac{1}{\lambda_j^*(1-tS_j)} - \frac{1}{\lambda_j^*(1-tS_{j-1})},$$

$$\begin{aligned} \frac{t \mathbb{E}D_j^q}{1-tS_j} &= \frac{t^2 S_K}{(1-tS_j)^2(1-tS_{j-1})}, \\ &= \frac{S_K}{\lambda_j^* S_j (1-tS_j)^2} + \frac{(\lambda_j^{*2} - S_j^2) S_K}{S_j S_{j-1} \lambda_j^{*2} (1-tS_j)} + \frac{S_K}{\lambda_j^{*2} (1-tS_{j-1})}, \end{aligned}$$

et,

$$\begin{aligned} \frac{t \mathbb{E}D_j^q}{1-tS_{j-1}} &= \frac{t^2 S_K}{(1-tS_j)(1-tS_{j-1})^2}, \\ &= -\frac{S_K}{\lambda_j^* S_{j-1} (1-tS_{j-1})^2} + \frac{(\lambda_j^{*2} - S_{j-1}^2) S_K}{S_j S_{j-1} \lambda_j^{*2} (1-tS_{j-1})} + \frac{S_K}{\lambda_j^{*2} (1-tS_j)}. \end{aligned}$$

Finalement, par une simple intégration des différents éléments simples, nous obtenons le résultat du théorème.  $\blacksquare$

Le prix optimal  $u_i^*$  déterminé par l'équation (4.7), correspond à la somme des moyennes uniformes des externalités induites par un utilisateur de classe  $i$  sur l'ensemble des classes de service. D'après la relation de demande (4.6), nous définissons

alors le coût total  $C_i(\underline{\lambda}^*)$  perçu par un utilisateur de classe  $i$  par :

$$\begin{aligned} C_i(\underline{\lambda}^*) &= u_i^* + v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda}^*t)dt, \\ &= \sum_{j=1}^K v_j \lambda_j^* \int_0^1 t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\underline{\lambda}^*t)dt + v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda}^*t)dt. \end{aligned}$$

La propriété importante de cette tarification est de partager le coût de congestion global  $L$  du système entre les classes de service, ce qui n'est pas vérifié pour le mécanisme du coût marginal proposé dans [81]. Pour cela, nous utilisons le mécanisme de partage de coût de Aumann-Shapley [108].

**Théorème 10** *Le mécanisme de tarification  $\underline{u}^*$  partage le coût de congestion total du système  $L$  entre les classes de service, c'est-à-dire :*

$$\sum_{i=1}^K \lambda_i^* C_i(\underline{\lambda}^*) = L(\underline{\lambda}^*).$$

**Preuve** Le coût total pour un utilisateur de classe  $i$  est défini par le relation de demande 4.6 par :

$$C_i(\underline{\lambda}) = u_i + v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda}t)dt.$$

Notre tarification optimale donnée par le théorème 9 induit un coût total

$$C_i(\underline{\lambda}^*) = \sum_{j=1}^K v_j \lambda_j^* \int_0^1 t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\underline{\lambda}^*t)dt + v_i \int_0^1 \mathbb{E}D_i(\underline{\lambda}^*t)dt.$$

En regroupant les deux intégrations nous obtenons :

$$\begin{aligned} C_i(\underline{\lambda}^*) &= \int_0^1 \left( \sum_{j=1}^K v_j \lambda_j^* t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(t\underline{\lambda}^*) + v_i \mathbb{E}D_i(t\underline{\lambda}^*) \right) dt, \\ &= \int_0^1 \frac{\partial \sum_{j=1}^K v_j \lambda_j^* \mathbb{E}D_j}{\partial \lambda_i}(t\underline{\lambda}^*) dt, \\ &= \int_0^1 \frac{\partial L}{\partial \lambda_i}(t\underline{\lambda}^*) dt. \end{aligned}$$

Le prix défini par  $\int_0^1 \frac{\partial L}{\partial \lambda_i}(t\underline{\lambda}^*) dt := C_i^{AS}(L, \underline{\lambda}^*)$  est connue dans la littérature économique sous le nom de mécanisme de Aumann-Shapley [108]. Ce mécanisme a été construit afin de répondre à 5 axiomes précis. Le premier axiome, appelé axiome de partage de coût, affirme que pour tout  $\underline{x} \in \mathbb{R}^n$ , nous avons

$$\sum_{i=1}^K x_i C_i^{AS}(F, \underline{x}) = F(\underline{x}).$$

Donc, en remarquant que notre tarification implique un coût par classe correspondant au mécanisme de Aumann-Shapley, nous obtenons que le mécanisme de tarification  $u^*$  défini par l'équation (4.7) partage le coût de congestion global  $L$  du système. ■

Nous avons construit un mécanisme de tarification qui force les utilisateurs à choisir le débit optimal au sens du surplus collectif et de plus qui partage le coût total du système entre les classes de service. Nous supposons maintenant que les utilisateurs ont le choix de leur classe de service (cas des classes ouvertes du chapitre précédent). Nous montrons que, dans le cas des temps de service homogènes, notre tarification incite les utilisateurs à choisir leur véritable classe de service. Cette propriété est connue sous le nom de compatibilité d'incitation et est souvent étudiée pour les mécanismes de tarification [24, 100].

### 4.2.2 Propriété de compatibilité d'incitation

Nous supposons que les utilisateurs et le système connaissent l'ensemble des caractéristiques  $\{(V_i, v_i)\}_{i=1,2,\dots,R}$  des classes de service. Par contre seul un utilisateur infinitésimal a connaissance de sa propre classe de service. En effet, nous utilisons la  $c - \mu$  rule qui implique que les classes de priorité sont ordonnées suivant l'ordre décroissant des valuations des délais, i.e.  $v_1 > v_2 > \dots > v_K$ . Un utilisateur dont les paquets appartiennent initialement à la classe  $i$  (dont la valuation du délai est  $v_i$ ), peut avoir un intérêt à choisir une classe de service  $j < i$  car il augmente le niveau de priorité de ses paquets. L'utilisateur peut également réduire le prix lorsque  $j > i$  car l'externalité, et donc le prix, augmente avec le niveau de priorité.<sup>1</sup> Ainsi, nous voulons que la décision des utilisateurs induite par la tarification force au bon choix de leur classe de service, c'est-à-dire au respect de la  $c - \mu$  rule, i.e. un utilisateur avec une valuation du délai  $v_i$  choisit la  $i^{\text{eme}}$  classe de service. Cette propriété est une autre interprétation de la compatibilité d'incitation que nous avons déjà rencontrée dans le chapitre 2 et est également étudiée pour un modèle de tarification d'un système multi-classe dans [87]. Dans les deux cas, nous voulons inciter les utilisateurs à réagir d'une certaine façon. Les utilisateurs choisissent leur classe de service suivant leur coût. Le coût  $C_i^j$  pour un utilisateur de classe  $i$  qui choisit la classe de service  $j$ , est composé suivant l'équation (4.6) d'un prix  $u_j^*$  et d'un coût en fonction au délai moyen  $\mathbb{E}D_j$  par

$$C_i^j(\underline{\lambda}^*) = u_j^* + v_i \int_0^1 \mathbb{E}D_j(\underline{\lambda}^*t) dt. \quad (4.9)$$

**Définition 4** *Un mécanisme de tarification  $\underline{u}$  possède la propriété de compatibilité d'incitation si pour toute classe  $i \in \{1, \dots, K\}$ ,*

$$i = \operatorname{argmin}_{j \in \{1, \dots, K\}} C_i^j(\underline{\lambda}^*).$$

*En d'autres termes,  $\forall j \neq i$  nous avons  $C_i^i(\underline{\lambda}^*) < C_i^j(\underline{\lambda}^*)$ .*

<sup>1</sup>Un paquet de classe 1 induit un coût de délai sur l'ensemble des classes du système alors qu'un paquet de classe  $K$  induit un coût de délai seulement sur sa propre classe.

Un modèle de tarification multi-classes possède cette propriété si les utilisateurs ont intérêt à choisir leur véritable classe de service. Le théorème suivant montre que notre tarification optimale  $\underline{u}^*$  défini par le théorème 9 possède cette propriété.

**Théorème 11** *Soit  $\underline{u}^*$  le vecteur des prix optimaux donné dans le théorème 9. Alors, ce mécanisme de tarification vérifie la propriété de compatibilité d'incitation.*

**Preuve** Considérons un job de classe  $i$  avec  $i \in \{1, \dots, K\}$ . Nous devons vérifier que pour tout  $k \in \{1, \dots, K\} \setminus \{i\}$

$$C_i^i(\underline{\lambda}^*) < C_i^k(\underline{\lambda}^*).$$

Or, nous avons

$$C_i^i(\underline{\lambda}^*) - C_i^k(\underline{\lambda}^*) = u_i^* - u_k^* + v_i \int_0^1 (\mathbb{E}D_i(\underline{\lambda}^*t) - \mathbb{E}D_k(\underline{\lambda}^*t)) dt,$$

où  $u_i^*$  et  $u_k^*$  sont définis dans le théorème 9. Ceci donne

$$\begin{aligned} C_i^i(\underline{\lambda}^*) - C_i^k(\underline{\lambda}^*) &= \int_0^1 \left( \sum_{j=1}^K v_j \lambda_j^* t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\underline{\lambda}^*t) \right) - \left( \sum_{j=1}^K v_j \lambda_j^* t \frac{\partial \mathbb{E}D_j}{\partial \lambda_k}(\underline{\lambda}^*t) \right) dt \\ &\quad + v_i \int_0^1 (\mathbb{E}D_i(\underline{\lambda}^*t) - \mathbb{E}D_k(\underline{\lambda}^*t)) dt, \\ &= \int_0^1 (\hat{u}_i(\underline{\lambda}^*t) - \hat{u}_k(\underline{\lambda}^*t) + v_i(\mathbb{E}D_i(\underline{\lambda}^*t) - \mathbb{E}D_k(\underline{\lambda}^*t))) dt, \end{aligned}$$

avec  $\hat{u}_i = \sum_{j=1}^K v_j \lambda_j^* t \frac{\partial \mathbb{E}D_j}{\partial \lambda_i}(\underline{\lambda}^*t)$ . Dans [81], Mendelson et Whang ont prouvé dans leur théorème 2, en considérant le mécanisme de coût marginal que,

$$\hat{u}_i(\underline{\lambda}^*) + v_i \mathbb{E}D_i(\underline{\lambda}^*) < \hat{u}_k(\underline{\lambda}^*) + v_i \mathbb{E}D_k(\underline{\lambda}^*),$$

pour tout  $k \in \{1, \dots, K\} \setminus \{i\}$ . Cette relation est également vérifiée en  $\underline{\lambda}^*t$ , i.e.

$$\hat{u}_i(\underline{\lambda}^*t) + v_i \mathbb{E}D_i(\underline{\lambda}^*t) < \hat{u}_k(\underline{\lambda}^*t) + v_i \mathbb{E}D_k(\underline{\lambda}^*t),$$

pour tout  $k \in \{1, \dots, K\} \setminus \{i\}$  et  $t \in [0, 1]$ . Donc, par une simple intégration d'une fonction positive, nous obtenons le résultat du théorème. ■

Ce théorème montre que, dans un contexte de temps de service homogène, le vecteur de prix  $\underline{u}^*$  construit dans le théorème 9 vérifie l'hypothèse de compatibilité d'incitation. De plus, par rapport au mécanisme du coût marginal, ce mécanisme partage le coût total du système entre les classes de service.

*Exemple* Considérons l'exemple proposé dans [81]. Il est composé d'un système avec deux classes de service et les fonctions de valuation suivantes

$$V_1(\lambda_1) = 9\lambda_1 - 10\lambda_1^2 \quad \text{et} \quad V_2(\lambda_2) = 12\lambda_2 - 15\lambda_2^2,$$

définies respectivement sur les intervalles  $[0, 9/20]$  et  $[0, 0.4]$ . Nous supposons les coûts de délai  $v_1 = 2$ ,  $v_2 = 1$  et des temps de service homogènes tels que  $\mu_1 = \mu_2 = 1$ . Les temps moyens d'attente pour chaque classe de service (donnés par l'équation (4.2)) sont

$$\begin{aligned} \mathbb{E}D_1(\underline{\lambda}) &= \frac{\lambda_1 + \lambda_2}{1 - \lambda_1} + 1, \\ \mathbb{E}D_2(\underline{\lambda}) &= \frac{\lambda_1 + \lambda_2}{(1 - \lambda_1)(1 - \lambda_1 - \lambda_2)} + 1. \end{aligned}$$

Nous obtenons les taux d'arrivées optimum  $\lambda_1^* = 0.183$  et  $\lambda_2^* = 0.278$  en résolvant le système obtenu avec les conditions du premier ordre du problème d'optimisation (4.4). Dans le tableau 4.1, nous donnons les coûts totaux et prix obtenus avec le mécanisme du coût marginal et le mécanisme décrit dans le théorème 9. Tout d'abord, nous vérifions la propriété de compatibilité d'incitation pour notre modèle de tarification. En effet, le coût global du système en terme de délai est égale à

$$v_1 \lambda_1^* \mathbb{E}D_1(\underline{\lambda}^*) + v_2 \lambda_2^* \mathbb{E}D_2(\underline{\lambda}^*) = 1.1415.$$

Nous avons

$$\lambda_1^* C_1^1 + \lambda_2^* C_2^2 = 1.1415,$$

et donc notre mécanisme partage le coût total entre les classes de service. Pour le mécanisme du coût marginal, nous obtenons

$$\lambda_1^* C_1^1 + \lambda_2^* C_2^2 = 1.99.$$

On remarque tout d'abord que le mécanisme du coût marginal sur-estime le coût total. Nous observons aussi que le coût pour chacune des classes est plus important que le coût avec notre mécanisme de tarification. Les utilisateurs vont donc préférer un fournisseur utilisant notre mécanisme de tarification car leur coût est inférieur à un autre fournisseur qui utilise une tarification avec coût marginal, avec une qualité de service (délai) égal.

	Coût marginal	Aumann-Shapley
$C_1^1$	5.34	3.2604
$C_1^2$	5.71	3.3514
$C_2^2$	3.66	1.9597
$C_2^1$	3.78	1.9976
$u_1^*$	2.21	0.7345
$u_2^*$	1.61	0.568

TAB. 4.1 – Coût total et prix pour chacune des classes de service.

### 4.3 Temps de service non-homogènes

Dans cette partie, nous étudions la propriété de compatibilité d'incitation pour notre prix optimal  $\underline{u}^*$  défini dans le théorème 9 avec des temps de service non-homogènes. En effet, dans un réseau IP comme l'Internet, les paquets à traiter dans les routeurs ne sont pas tous de même longueur et ne demandent ainsi pas le même temps de traitement. Tout d'abord, nous considérons l'exemple précédent avec des taux de service différents pour chacune des classes.

*Exemple* Reprenons l'exemple précédent avec les temps moyen de service  $1/\mu_1 = 0.1$  et  $1/\mu_2 = 2$ . L'optimisation du surplus collectif donne les taux d'arrivées optimaux  $\lambda_1^* = 0.37$  et  $\lambda_2^* = 0.15$ . En utilisant le théorème 9 nous obtenons les prix optimaux  $u_1^* = 0.52$  et  $u_2^* = 2.76$ . Ainsi, nous obtenons les coûts globaux suivants :

$$C_1^1(\underline{\lambda}^*) = 1.53 \quad C_1^2(\underline{\lambda}^*) = 8.07$$

et

$$C_2^2(\underline{\lambda}^*) = 5.42 \quad C_2^1(\underline{\lambda}^*) = 1.03.$$

Nous pouvons alors déduire que les utilisateurs de classe 2 admettent un coût moins important en se faisant traiter dans la première classe de service.

Cet exemple montre que le mécanisme de tarification  $\underline{u}^*$  construit dans le théorème 9 ne vérifie pas la propriété de compatibilité d'incitation lorsque les temps de service sont non-homogènes. Afin de résoudre ce problème, nous imposons que le prix  $u_i$  pour un utilisateur de classe  $i$  soit fonction du temps de service  $\tau_i$ . Définissons par  $\mathbb{E}(u_i(\tau_i))$  le prix moyen sur la période de service de taux  $\tau_i$ . A partir de l'expression (4.6), nous considérons la nouvelle relation de demande

$$\frac{V_i(\lambda_i)}{\lambda_i} = \mathbb{E}(u_i(\tau_i)) + v_i \int_0^1 \mathbb{E}D_i(\lambda t) dt. \quad (4.10)$$

Nous reprenons en nous basant sur [81], une fonction de prix quadratique par rapport aux temps de services et cherchons les paramètres de cette fonction. Le théorème suivant donne ces prix optimaux et montre que cette nouvelle tarification vérifie la propriété de compatibilité d'incitation.

**Proposition 5** Soit  $u_i^*(\tau_i)$  défini par

$$u_i^*(\tau_i) = A_i' \tau_i + 1/2 B_i' \tau_i^2, \quad (4.11)$$

avec

$$\begin{aligned}
A'_i(\underline{\lambda}) &= \left( \sum_{j=1}^K c_j^2 \lambda_j^* / \mu_j^2 \right) \left[ \sum_{k=i}^K v_k \left( \frac{1}{c_k S_k (1 - S_k)} \right. \right. \\
&\quad \left. \left. - \frac{(c_k^2 \lambda_k^{*2} - S_k^2) \ln(1 - S_k)}{S_k^2 S_{k-1} c_k^2 \lambda_k^*} - \frac{\ln(1 - S_{k-1})}{c_k^2 \lambda_k^* S_{k-1}} \right) \right. \\
&\quad \left. + \sum_{k=i+1}^K v_k \left( \frac{1}{c_k S_{k-1} (1 - S_{k-1})} - \frac{(c_k^2 \lambda_k^{*2} - S_{k-1}^2) \ln(1 - S_{k-1})}{S_k S_{k-1} c_k^2 \lambda_k^*} \right. \right. \\
&\quad \left. \left. - \frac{\ln(1 - S_k)}{c_k^2 \lambda_k^* S_k} \right) \right]
\end{aligned}$$

et

$$B'(\underline{\lambda}) = - \sum_{k=1}^K v_k \left( \frac{\ln(1 - S_{k-1})}{S_{k-1}} + \frac{\ln(1 - S_k)}{S_k} \right),$$

où  $S_k = \sum_{i=1}^k c_k \lambda_k^*$  et  $c_j = 1/\mu_j$ . Le modèle de tarification basé sur cette fonction de prix est optimal et vérifie la propriété de compatibilité d'incitation et est optimal.

**Preuve** Dans [81], les auteurs ont montré dans le cas du coût marginal (avec une autre fonction de demande), que le mécanisme de tarification  $\hat{u}_i^*(\underline{\lambda}, \tau_i) = A_i(\underline{\lambda})\tau_i + 1/2B(\underline{\lambda})\tau_i^2$  est optimale et vérifie la propriété de compatibilité d'incitation dans le cas de temps de service non-homogènes. Ils définissent les grandeurs suivantes

$$\begin{aligned}
A_i(\underline{\lambda}) &= \frac{a_i}{\bar{S}_{i-1} \bar{S}_i} + \sum_{k=i+1}^I a_k \left( \frac{1}{\bar{S}_{k-1}^2 \bar{S}_k} + \frac{1}{\bar{S}_{k-1} \bar{S}_k^2} \right), \\
&= \sum_{k=i}^K v_k \lambda_k^* \left( \sum_{j=1}^K c_j^2 \lambda_j^{*2} \frac{1}{(1 - S_{k-1})(1 - S_k)^2} \right) \\
&\quad + \sum_{k=i+1}^K v_k \lambda_k^* \left( \sum_{j=1}^K c_j^2 \lambda_j^{*2} \frac{1}{(1 - S_{k-1})^2 (1 - S_k)} \right)
\end{aligned}$$

et

$$B(\underline{\lambda}) = \sum_{k=1}^K \frac{v_k \lambda_k^*}{(1 - S_{k-1})(1 - S_k)},$$

où  $a_i = v_i \lambda_i^* \sum_{k=1}^K c_k^2 \lambda_k^*$  et  $c_j = 1/\mu_j$ . Par le biais d'une intégration, définissons

$$A'_i = \int_0^1 A_i(u\underline{\lambda}) du \quad \text{et} \quad B' = \int_0^1 B(u\underline{\lambda}) du.$$

Nous montrons que cette tarification basée sur la fonction de prix

$$u_i^*(\tau_i) = A_i' \tau_i + 1/2 B' \tau_i^2$$

vérifie la propriété de compatibilité d'incitation et est optimale, en utilisant les résultats sur le mécanisme du coût marginal.

Le mécanisme de tarification  $u^*$  défini par l'équation (4.11) est optimal car

$$\mathbb{E}(u_i^*(\tau_i)) = u_i^*, \quad \forall i \in \{1, \dots, K\},$$

où  $u_i^*$  est le prix optimal donné par le théorème 9. En effet, en moyenne, les taux d'arrivées qui maximisent le surplus collectif du système sont les  $\lambda^*$  déterminés par l'optimisation (4.4). A partir des définitions de  $A_i'$  et  $B'$ , nous observons que  $u_i^*(t)$  est la moyenne uniforme du mécanisme basé sur la fonction de prix  $\hat{u}_i^*$  de Mendelson et Whang [81, page 879] définie par

$$u_i^*(t) = \int_0^1 \hat{u}_i^*(u\lambda, t) du.$$

Donc, nous obtenons

$$\begin{aligned} \mathbb{E}(u_i^*(\tau_i)) &= \mathbb{E} \left( \int_0^1 \hat{u}_i^*(u\lambda, \tau_i) du \right), \\ &= \int_0^1 \mathbb{E}(\hat{u}_i^*(u\lambda, \tau_i)) du, \\ &= \int_0^1 \hat{u}_i^*(u\lambda) du, \quad (\text{depuis [81, page 879]}) \\ &= u_i^*. \end{aligned}$$

Le mécanisme est donc optimal car la moyenne du prix sur une période pour un utilisateur de classe  $i$  est le prix optimal  $u_i^*$ .

De plus, le mécanisme du coût marginal basé sur la fonction de prix  $\hat{u}_i^*(\lambda, t)$  vérifie la propriété de compatibilité d'incitation, c'est-à-dire

$$\forall i, j \in \{1, \dots, K\}, \quad \hat{u}_i^*(\lambda, t) + v_i \mathbb{E} D_i(\lambda) \leq \hat{u}_j^*(\lambda, t) + v_i \mathbb{E} D_j(\lambda),$$

pour tout  $t \geq 0$ . Cette relation est également vérifiée en  $\lambda u$ , i.e.

$$\hat{u}_i^*(\lambda u, t) + v_i \mathbb{E} D_i(\lambda u) \leq \hat{u}_j^*(\lambda u, t) + v_i \mathbb{E} D_j(\lambda u),$$

pour tout  $j \in \{1, \dots, n\} \setminus \{i\}$ ,  $t \geq 0$  et  $u \in [0, 1]$ .

Par conséquent, pour tout  $j \in \{1, \dots, n\} \setminus \{i\}$  et  $t \geq 0$  nous avons :

$$\begin{aligned} \int_0^1 (\hat{u}_i^*(u\lambda, t) + v_i \mathbb{E} D_i(u\lambda)) du &\leq \int_0^1 (\hat{u}_j^*(u\lambda, t) + v_i \mathbb{E} D_j(u\lambda)) du, \\ u_i^*(t) + v_i \int_0^1 \mathbb{E} D_i(u\lambda) du &\leq u_j^*(t) + v_i \int_0^1 \mathbb{E} D_j(u\lambda) du, \end{aligned}$$

ce qui montre que le nouveau modèle basé sur la fonction de prix  $\hat{p}_i^*(\tau_i)$  vérifie la propriété de *Compatibilité d'Incitation*. ■



Dans la suite, nous étendons notre modèle à deux cas de figures. Tout d'abord, nous proposons un modèle de tarification dynamique où les taux d'arrivées varient dans le temps suivant le prix et les débits obtenus dans le passé. Enfin, nous considérons l'extension de notre mécanisme à un réseau ouvert de files d'attentes avec modélisation du routage des paquets entre les files.

## 4.4 Tarification dynamique mono-classe

Nous considérons un modèle de tarification dynamique où les taux d'arrivées varient d'une période de temps à l'autre. Ces mécanismes sont plus efficaces par rapport à la tarification statique car ils réagissent à la moindre variabilité de la demande, néanmoins ils possèdent de nombreux inconvénients comme la signalisation (gestion permanente des données de mis-à-jour à travers le réseau) ou encore la granularité (découpage du temps) [89]. Malgré ces inconvénients, nous étudions notre mécanisme de partage de coût de congestion optimal dans un contexte dynamique. Nous supposons une unique classe ( $K = 1$ ) de service et sans perte de généralités, que le taux de service est  $\mu = 1$ . La généralisation au cas multi-classes n'engendre pas de difficultés particulières de modélisation mais des expressions analytiques plus complexes. Le temps est découpé en périodes durant lesquelles les taux d'arrivées sont supposés constants. Nous notons  $\lambda_t$  le taux d'arrivée des paquets pendant la période  $t$ . A la fin de chaque période  $t$ , le prix pour la période suivante  $t + 1$  est calculé en fonction du taux d'arrivée  $\lambda_t$  à partir du prix optimal déterminé dans le théorème 9. Nous pouvons écrire la relation suivante à partir de l'équation (4.7)

$$u_{t+1}^+ = u(\lambda_t) = v\lambda_t \int_0^1 x \mathbb{E}D'(x\lambda_t) dx. \quad (4.12)$$

Nous supposons que les taux d'arrivées ont un comportement "Markovien" dans le temps, c'est-à-dire que le taux  $\lambda_{t+1}$  est fonction seulement du taux d'arrivée de la période précédente  $\lambda_t$  induit par la relation de demande

$$\frac{V(\lambda_{t+1})}{\lambda_{t+1}} = u_{t+1}^+ + v \int_0^1 \mathbb{E}D(x\lambda_t) dx = f(\lambda_t). \quad (4.13)$$

*Exemple* Nous considérons l'exemple d'un système modélisé par une file d'attente de type M/M/1 avec la fonction de valuation  $V(\lambda) = 5\lambda - 4\lambda^2$  et  $\mu = 1$ . Pour s'assurer que le taux d'arrivée des paquets ne soit pas négatif ou supérieur au taux de service  $\mu = 1$ , le système contrôle le débit des utilisateurs; ainsi nous posons pour  $\epsilon > 0$  et pour toutes périodes  $t$

$$\tilde{f}(\lambda_t) = \begin{cases} \epsilon & \text{si } f(\lambda_t) < 0, \\ 1 - \epsilon & \text{si } f(\lambda_t) > 1 - \epsilon, \\ f(\lambda_t) & \text{sinon.} \end{cases}$$

Nous considérons la valeur  $\epsilon = 0.05$ . A partir de la relation de demande( 4.13) et la fonction de prix (4.12), nous obtenons l'expression du taux d'arrivée  $\lambda_{t+1}$  en fonction

du taux d'arrivée  $\lambda_t$  par

$$V(\lambda_{t+1}) = \lambda_{t+1} \int_0^1 (\lambda_t x \mathbb{E}D'(x\lambda_t) + \mathbb{E}D(x\lambda_t)) dx, \quad (4.14)$$

avec  $\mathbb{E}D(\lambda) = \frac{1}{1-\lambda}$ , ce qui donne

$$V(\lambda_{t+1}) = \lambda_{t+1} \mathbb{E}D(\lambda_t) = \lambda_{t+1} \frac{1}{1-\lambda_t},$$

puis

$$\lambda_{t+1} = \frac{1}{4} \left( 5 - \frac{1}{1-\lambda_t} \right) = f(\lambda_t).$$

La figure 4.1 illustre l'instabilité du taux d'arrivée dans le temps et par voie de conséquence l'instabilité du prix.

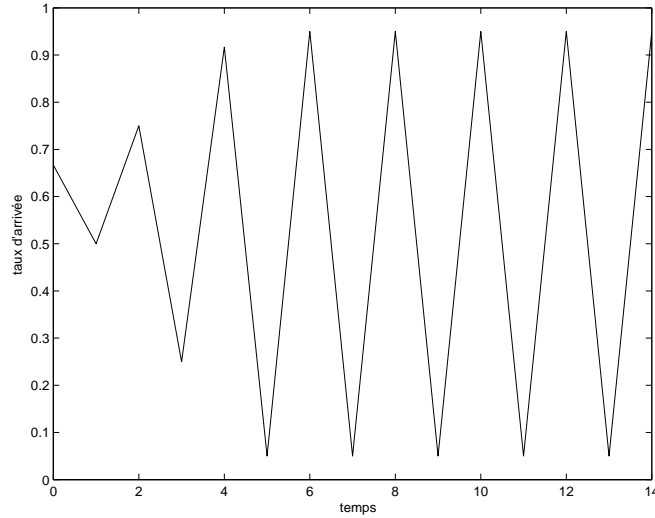


FIG. 4.1 – Taux d'arrivée divergent en fonction du temps pour le modèle de tarification dynamique "Markovien".

Le théorème suivant donne une condition nécessaire et suffisante sur les paramètres du modèle afin de prévenir ce comportement oscillatoire des prix et taux d'arrivées.

**Théorème 12** *Nous considérons un système modélisé par une file d'attente M/M/1 avec un taux de service  $\mu = 1$  où les taux d'arrivées évoluent dans le temps de façon Markovien. Le système est stable, i.e. le taux d'arrivée converge, si et seulement si*

$$\frac{v}{(1-\lambda^*)^2} < \frac{|v(1-\lambda^*)\lambda^*V'(v\frac{\lambda^*}{1-\lambda^*}) - (1-\lambda^*)^2V(v\frac{\lambda^*}{1-\lambda^*})|}{v^2\lambda^{*2}}, \quad (4.15)$$

où  $\lambda^*$  est la solution de l'équation du point-fixe  $\lambda = f(\lambda)$  avec  $f$  définie en (4.13).

On remarquera que si le système est stable, en régime stationnaire, le taux d'arrivée est égal à  $\lambda^*$ , solution de l'équation  $f(\lambda) = \lambda$ .

**Preuve** À partir de la relation de demande (4.13), définissons la fonction

$$\begin{aligned} U(\lambda_{t+1}) &= \frac{V(\lambda_{t+1})}{\lambda_{t+1}} = u + v \int_0^1 \mathbb{E}D(x[\lambda_{t+1}|\bar{\lambda}_t])dx, \\ &= u + v \int_0^1 \mathbb{E}D(x\lambda_t)dx, \\ &= g(\lambda_t), \end{aligned} \quad (4.16)$$

de sorte que nous avons

$$\lambda_{t+1} = U^{-1}(g(\lambda_t)) = f(\lambda_t).$$

Une condition nécessaire et suffisante de convergence est  $|f'(\lambda^*)| < 1$ , avec  $\lambda^*$  solution de l'équation  $\lambda^* = f(\lambda^*)$ . De plus, si il y a convergence, le taux d'arrivée converge vers  $\lambda^*$ . La dérivée première de la fonction  $f$  est

$$\begin{aligned} f'(\lambda^*) &= (U^{-1} \circ g)'(\lambda^*), \\ &= (U^{-1})'(g(\lambda^*))g'(\lambda^*). \end{aligned}$$

En utilisant les équations (4.12) et (4.16) nous obtenons

$$g(\lambda^*) = v\lambda^* \int_0^1 \frac{t}{(1-t\lambda^*)^2} dt + v\lambda^* \int_0^1 \frac{t}{1-t\lambda^*} dt = v \frac{\lambda^*}{1-\lambda^*},$$

et de plus,

$$g'(\lambda^*) = \frac{v}{(1-\lambda^*)^2}.$$

Nous avons la relation suivante entre les fonctions  $U$  et  $V$  :

$$U'(g(\lambda^*)) = \frac{v(1-\lambda^*)\lambda^*V'(v\frac{\lambda^*}{1-\lambda^*}) - (1-\lambda^*)^2V(v\frac{\lambda^*}{1-\lambda^*})}{v^2\lambda^{*2}},$$

car  $U'(\lambda_{t+1}) = \frac{V(\lambda_{t+1})}{\lambda_{t+1}} - \frac{V(\lambda_{t+1})}{\lambda_{t+1}^2}$ . Ainsi, la condition  $|f'(\lambda^*)| < 1$  est équivalente à  $|g'(\lambda^*)| < |U'(g(\lambda^*))|$ . ■

Afin d'illustrer ce théorème, nous reprenons l'exemple précédent.

*Exemple (suite)* Le taux d'arrivée tel que  $f(\lambda) = \lambda$  est  $\lambda^* = 0.6096$ . Nous avons  $|g'(\lambda^*)| = \frac{1}{(1-0.64)^2} = 6.56$  et  $|U'(g(\lambda^*))| = 4$  ce qui implique que la condition de stabilité donnée dans le théorème 12 n'est pas vérifiée. Ceci explique le caractère oscillatoire du taux d'arrivée.

Changeons la fonction de valuation par  $V(\lambda) = 5\lambda - 10\lambda^2$  et conservons les autres paramètres. Dans ce cas, la solution du point fixe est  $\lambda^* = 0.4285$  et nous obtenons  $|g'(\lambda^*)| = \frac{1}{(1-0.42)^2} = 2.97$  et  $|U'(g(\lambda^*))| = 10$ . Ainsi, pour ce nouvel exemple, la condition de stabilité donnée par le théorème 12 est vérifiée. La figure 4.2 représente l'évolution du taux d'arrivée en fonction du temps et montre sa convergence vers la solution du point fixe  $\lambda^* = 0.4285$ .

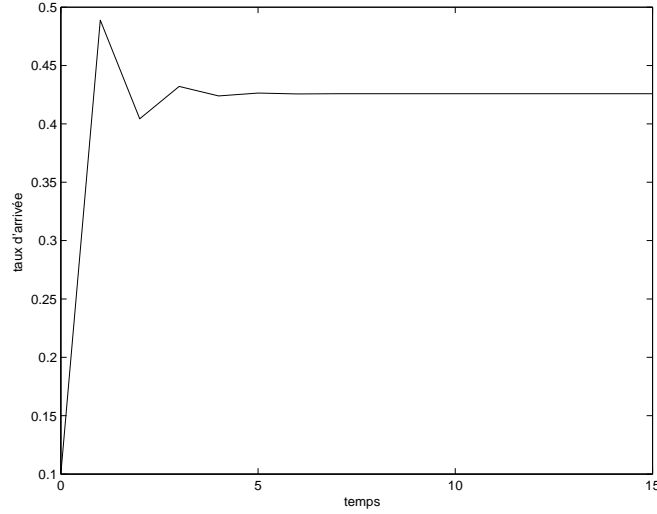


FIG. 4.2 – Taux d’arrivée convergent en fonction du temps pour le modèle de tarification dynamique ”Markovien”.

## 4.5 Extension à un réseau de Jackson ouvert

Nous nous intéressons maintenant à l’extension de notre mécanisme de tarification dans un réseau ouvert de Jackson [64]. Ce modèle de réseau provient de [78] pour lequel les auteurs utilisent la tarification au coût marginal. Nous considérons un réseau ouvert composé de  $M$  nœuds indexés par  $k \in \{1, 2, \dots, M\}$ . Les paquets entrent dans le réseau au nœud  $k$  avec un taux  $\Lambda_k$ . Les arrivées des paquets de classe  $i$ , pour  $i \in \{1, \dots, K\}$ , suivent un processus de Poisson de taux  $\lambda^i$  et le vecteur des taux d’arrivées est  $\underline{\Lambda} = (\Lambda^1, \dots, \Lambda^K)$ . Chaque paquet possède une route à travers le réseau notée  $r$  qui représente la série de nœuds traversés. La politique de service est PAPS pour chaque nœud. Le temps de service au nœud  $k$  suit une loi exponentielle de taux  $\mu_k$  et est indépendant du taux d’arrivée des paquets. Nous notons  $Q^i$  la matrice de routage du réseau où  $Q_{k',k}^i$  est la probabilité qu’un paquet de classe  $i$  transite du nœud  $k'$  vers le nœud  $k$ . La route d’un paquet de classe  $i$  est notée  $r^i$ . Le nœud de destination d’un paquet de classe  $i$  est un nœud absorbant de  $Q^i$ . De plus, nous notons  $\underline{q}^i$  le vecteur où  $q_k^i$  est la probabilité qu’un paquet de classe  $i$  entre dans le réseau par le nœud  $k$ . Nous posons  $\mathbb{E}D_k$  le temps moyen d’attente au nœud  $k$  et le coût de congestion moyen pour un paquet de classe  $i$  est la moyenne sur les routes  $v\mathbb{E}D^i = v\mathbb{E}(\sum_{k \in r^i} \mathbb{E}D_k)$ . Nous notons par  $\lambda_k^i$  le taux d’arrivée effectif des paquets de classe  $i$  au nœud  $k$ . Le vecteur  $\underline{\lambda}^i$  est solution unique du système linéaire suivant :

$$\begin{aligned} \underline{\lambda}^i &= \underline{\lambda}^i Q^i + \Lambda^i \underline{q}^i, \\ \underline{\lambda}^i &= \Lambda^i \underline{b}^i, \end{aligned} \tag{4.17}$$

avec  $\underline{b}^i = \underline{q}^i(I - Q^i)^{-1}$ ,  $\underline{\lambda}^i = (\lambda_1^i, \lambda_2^i, \dots, \lambda_M^i)$  et  $\underline{q}^i = (q_1^i, q_2^i, \dots, q_M^i)$  avec  $I$  la matrice identité.

Soit  $u^i$  le prix unitaire pour un paquet de classe  $i$ . A partir de la relation 4.6, le taux d'arrivée des paquets est défini par la relation de demande suivante

$$\int_0^1 (V^i)'(\Lambda^i t) dt = \frac{V^i(\Lambda^i)}{\Lambda^i} = u^i + v \int_0^1 \mathbb{E}D^i(\underline{\Lambda}t) dt. \quad (4.18)$$

Nous avons la proposition suivante qui donne le prix optimal en prenant en compte la relation de demande 4.18.

**Proposition 6** *Le prix optimal  $u^i$  pour un paquet de classe  $i$  est*

$$u^i(\underline{\Lambda}^*) = \sum_{k \in r^i} b_k^i u_k(\underline{\Lambda}^*), \quad (4.19)$$

où  $\underline{\Lambda}^*$  est le vecteur des taux d'arrivées qui maximisent le surplus collectif  $\sum_i (V^i(\Lambda^i) - v\Lambda^i \mathbb{E}D^i(\underline{\Lambda}))$ . Le prix optimal au nœud  $k$  est

$$u_k(\underline{\Lambda}^*) = v \sum_i (b_k^i \Lambda^{i*}) \int_0^1 t \mathbb{E}D'_k \left( t \sum_i b_k^i \Lambda^{i*} \right) dt.$$

De plus, le nœud  $k$  est une file d'attente de type  $M/M/1$  avec taux de service  $\mu_k$ , nous avons donc en régime stationnaire

$$u_k(\underline{\Lambda}^*) = \frac{v \ln \left( \frac{\mu_k - \sum_i (b_k^i \Lambda^{i*})}{\mu_k} \right)}{\sum_i b_k^i \Lambda^{i*}} + \frac{v}{\mu_k - \sum_i b_k^i \Lambda^{i*}}. \quad (4.20)$$

**Preuve** L'optimisation du surplus collectif s'écrit :

$$\begin{aligned} \max_{\underline{\Lambda}} NV(\underline{\Lambda}) &= \max_{\underline{\Lambda}} \left( \sum_i V^i(\Lambda^i) - \sum_i v \Lambda^i \mathbb{E}D^i(\underline{\Lambda}) \right), \\ &= \max_{\underline{\Lambda}} \left( \sum_i V^i(\Lambda^i) - \sum_k v \left( \sum_i b_k^i \Lambda^i \right) \mathbb{E}D_k \left( \sum_i b_k^i \Lambda^i \right) \right), \end{aligned} \quad (4.21)$$

où nous avons

$$\mathbb{E}D^i = \mathbb{E} \left( \sum_{k \in r^i} \mathbb{E}D_k \right) = \sum_{k \in r^i} b_k^i \mathbb{E}D_k \quad (4.22)$$

et

$$\mathbb{E}D_k(\underline{\lambda}_1, \underline{\lambda}_2, \dots, \underline{\lambda}_K) = \mathbb{E}D_k \left( \sum_{i=1}^I \lambda_k^i \right), \quad (4.23)$$

car  $b_k^i$  est le nombre moyen de visites au nœud  $k$  par un paquet de classe  $i$ . La condition du premier ordre pour l'optimisation du surplus collectif s'écrit :

$$\begin{aligned}
(V^i)'(\Lambda^{i*}) &= v\mathbb{E}D^i + v \sum_{j=1}^K \Lambda^j \frac{\partial \mathbb{E}D^j}{\partial \Lambda^i}, \\
&= v \sum_{k \in r^i} b_k^i \mathbb{E}D_k \left( \sum_{i=1}^K b_k^i \Lambda^{i*} \right) + v \sum_{j=1}^K \Lambda^j \frac{\partial}{\partial \Lambda^i} \left( \sum_{k \in r^j} b_k^j \mathbb{E}D_k \left( \sum_i b_k^i \Lambda^{i*} \right) \right), \\
&= v \sum_{k \in r^i} b_k^i \mathbb{E}D_k \left( \sum_{i=1}^K b_k^i \Lambda^{i*} \right) + v \sum_{j=1}^K \Lambda^j \sum_{k \in r^j} b_k^j \frac{\partial \mathbb{E}D_k}{\partial \Lambda^i} \left( \sum_i b_k^i \Lambda^{i*} \right), \\
&= v \sum_{k \in r^i} b_k^i \mathbb{E}D_k \left( \sum_{i=1}^K b_k^i \Lambda^{i*} \right) + v \sum_{j=1}^K \left( \sum_{k \in r^j} b_k^j \Lambda^j \right) \frac{\partial \mathbb{E}D_k}{\partial \Lambda^i} \left( \sum_i b_k^i \Lambda^{i*} \right), \\
&= v \sum_{k \in r^i} b_k^i \mathbb{E}D_k \left( \sum_{i=1}^K b_k^i \Lambda^{i*} \right) + \sum_k v \left( \sum_i b_k^i \Lambda^{i*} \right) b_k^i \mathbb{E}D'_k \left( \sum_i b_k^i \Lambda^{i*} \right).
\end{aligned}$$

À partir de la relation de demande (4.18), nous obtenons alors le prix optimal

$$\begin{aligned}
u^i(\underline{\Lambda}^*) &= \int_0^1 \sum_k v \left( \sum_i b_k^i \Lambda^{i*} t \right) b_k^i \mathbb{E}D'_k \left( t \sum_i b_k^i \Lambda^{i*} \right) dt, \\
&= \sum_k b_k^i v \left( \sum_i b_k^i \Lambda^{i*} \right) \int_0^1 t \mathbb{E}D'_k \left( t \sum_i b_k^i \Lambda^{i*} \right) dt, \\
&= \sum_k b_k^i u_k(\underline{\Lambda}^*).
\end{aligned}$$

Nous avons pour une file M/M/1 avec taux de service  $\mu$  :

$$\mathbb{E}D(\lambda) = \frac{1}{\mu - \lambda},$$

et

$$\mathbb{E}D'(\lambda) = \frac{1}{(\mu - \lambda)^2}.$$

Dans notre réseau de Jackson, le nœud  $k$  se comporte comme une file d'attente de type M/M/1 avec les taux d'arrivées  $\sum_{i=1}^M \lambda_k^i = \sum_i (b_k^i \Lambda^i)$  et taux de service  $\mu_k$ . Nous avons donc :

$$\mathbb{E}D'_k \left( t \sum_i (b_k^i \Lambda^{i*}) \right) = \frac{1}{(\mu_k - t \sum_i (b_k^i \Lambda^{i*}))^2},$$

et nous obtenons le prix optimal

$$u_k(\underline{\Lambda}^*) = \frac{v \ln \left( \frac{\mu_k - \sum_i (b_k^i \Lambda^{i*})}{\mu_k} \right)}{\sum_i b_k^i \Lambda^{i*}} + \frac{v}{\mu_k - \sum_i b_k^i \Lambda^{i*}}.$$

■

*Exemple* Nous considérons un exemple de réseau ouvert de Jackson représenté sur la figure 4.3. Ce réseau comporte quatre files d'attente et deux classes de trafic. Les routes empruntées par le trafic de classe 1 est en pointillé et pour le trafic de classe 2 en trait plein.

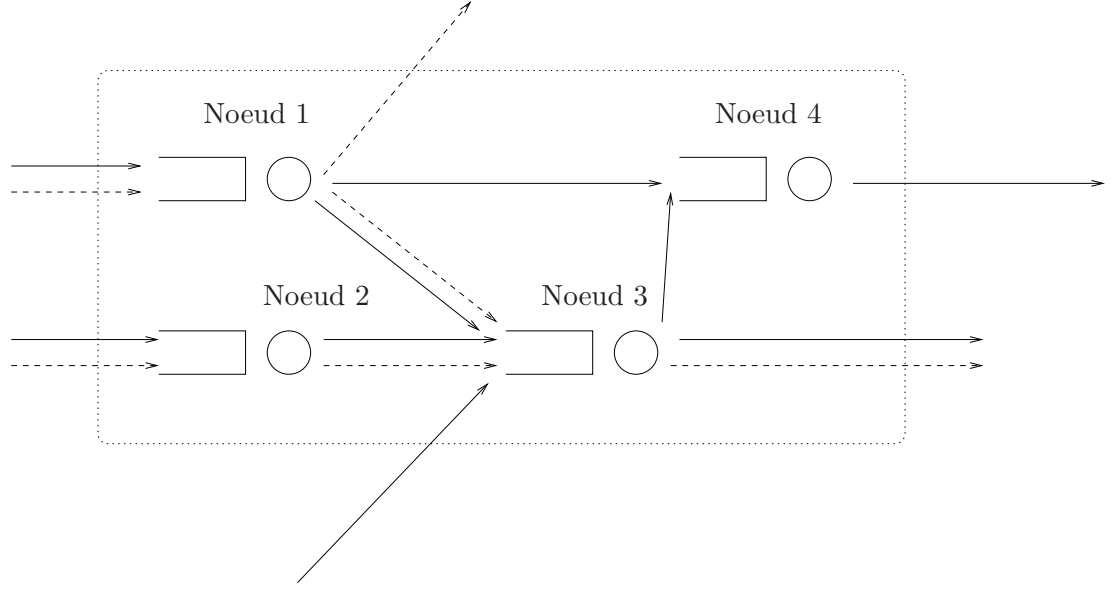


FIG. 4.3 – Exemple de réseau de Jackson avec deux types de trafic.

Nous avons  $M = 4$  files d'attente et  $K = 2$  types de trafic dans ce réseau. On considère que les taux de service de chaque station  $\mu_k$ , pour  $k = 1, \dots, 4$  sont égaux à 1 et également le paramètre de coût  $v = 1$ . Les fonctions de valuation pour chaque type de trafic sont :

$$V_1(\Lambda_1) = 5\Lambda_1 - \Lambda_1^2, \quad \text{et} \quad V_2(\Lambda_2) = 6\Lambda_2 - 3\Lambda_2^2.$$

Nous considérons les matrices de routages :

$$Q^1 = \begin{pmatrix} 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{et} \quad Q^2 = \begin{pmatrix} 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1/2 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

De plus, les vecteurs définissant les probabilités qu'un paquet de classe  $i$  joigne le réseau par le nœud  $k$  sont :

$$\underline{q}^1 = (1/2, 1/2, 0, 0) \quad \text{et} \quad \underline{q}^2 = (1/4, 1/4, 1/2, 0).$$

Les vecteurs du nombre moyens de visite à chaque station pour chaque type de trafic, donnés par la relation 4.17, sont :

$$\underline{b}^1 = \underline{q}^1(I - Q^1)^{-1} = (1/2, 1/2, 3/4, 0),$$

et

$$\underline{b}^2 = \underline{q}^2(I - Q^2)^{-1} = (1/4, 1/4, 15/16, 17/32).$$

Nous calculons le vecteur des débits  $\underline{\Lambda}^*$  qui optimise le surplus collectif du système  $NV(\underline{\Lambda}) = \sum_i (V^i(\Lambda^i) - v\Lambda^i \mathbb{E}D^i(\underline{\Lambda}))$ . Pour cet exemple, la fonction du surplus collectif à optimiser s'écrit plus simplement sous la forme :

$$NV(\Lambda^1, \Lambda^2) = V^1(\Lambda^1) + V^2(\Lambda^2) - \Lambda^1 \mathbb{E}D^1(\underline{\Lambda}) - \Lambda^2 \mathbb{E}D^2(\underline{\Lambda}),$$

avec, d'après (4.22),

$$\mathbb{E}D^1(\underline{\Lambda}) = \frac{1}{2}\mathbb{E}D_1(\underline{\Lambda}) + \frac{1}{2}\mathbb{E}D_2(\underline{\Lambda}) + \frac{3}{4}\mathbb{E}D_3(\underline{\Lambda}),$$

et

$$\mathbb{E}D^2(\underline{\Lambda}) = \frac{1}{4}\mathbb{E}D_1(\underline{\Lambda}) + \frac{1}{4}\mathbb{E}D_2(\underline{\Lambda}) + \frac{15}{16}\mathbb{E}D_3(\underline{\Lambda}) + \frac{17}{32}\mathbb{E}D_4(\underline{\Lambda}).$$

De plus, dans un réseau de Jackson, chaque file évolue comme une M/M/1 et donc les délais moyen d'un paquet à chacune des files s'expriment par :

$$\begin{aligned} \mathbb{E}D_1(\underline{\Lambda}) &= \frac{1}{1 - \frac{1}{2}\Lambda^1 - \frac{1}{4}\Lambda^2}, & \mathbb{E}D_2(\underline{\Lambda}) &= \frac{1}{1 - \frac{1}{2}\Lambda^1 - \frac{1}{4}\Lambda^2}, \\ \mathbb{E}D_3(\underline{\Lambda}) &= \frac{1}{1 - \frac{3}{4}\Lambda^1 - \frac{15}{16}\Lambda^2}, & \text{et } \mathbb{E}D_4(\underline{\Lambda}) &= \frac{1}{1 - \frac{17}{32}\Lambda^2}. \end{aligned}$$

Les débits optimaux  $(\Lambda^{1*}, \Lambda^{2*})$  sont solution du système suivant formé par les équations du premier ordre pour l'optimisation du surplus collectif qui sont :

$$(S) : \begin{cases} (V^1)(\Lambda^{1*})' = \mathbb{E}D^1(\Lambda^{1*}, \Lambda^{2*}) + \frac{\partial \mathbb{E}D^1}{\partial \Lambda_1}(\Lambda^{1*}, \Lambda^{2*}) + \frac{\partial \mathbb{E}D^2}{\partial \Lambda_1}(\Lambda^{1*}, \Lambda^{2*}) \\ (V^2)(\Lambda^{2*})' = \mathbb{E}D^2(\Lambda^{1*}, \Lambda^{2*}) + \frac{\partial \mathbb{E}D^1}{\partial \Lambda_2}(\Lambda^{1*}, \Lambda^{2*}) + \frac{\partial \mathbb{E}D^2}{\partial \Lambda_2}(\Lambda^{1*}, \Lambda^{2*}) \end{cases}$$

Or d'après les expressions des fonctions de valuation, nous avons

$$(V^1)(\Lambda^{1*})' = 5 - 2\Lambda^{1*} \quad \text{et} \quad (V^2)(\Lambda^{2*})' = 6 - 6\Lambda^{2*}.$$

De plus, d'après les délais moyen d'un paquet à chaque file, le système (S) s'écrit après quelques simplifications d'écriture :

$$(S) : \begin{cases} 5 - 2\Lambda^{1*} = \frac{1}{(1 - \frac{1}{2}\Lambda^{1*} - \frac{1}{4}\Lambda^{2*})^2} + \frac{3/4}{(1 - \frac{3}{4}\Lambda^{1*} - \frac{15}{16}\Lambda^{2*})^2} \\ 6 - 6\Lambda^{2*} = \frac{1/2}{(1 - \frac{1}{2}\Lambda^{1*} - \frac{1}{4}\Lambda^{2*})^2} + \frac{15/16}{(1 - \frac{3}{4}\Lambda^{1*} - \frac{15}{16}\Lambda^{2*})^2} + \frac{17/32}{(1 - \frac{17}{32}\Lambda^{2*})^2} \end{cases}$$

Ce système possède une unique solution réelle

$$(\Lambda^{1*}, \Lambda^{2*}) = (0.357, 0.206).$$

Nous pouvons maintenant déterminer, d'après le théorème 6, les prix optimaux pour chaque nœud du réseau tout d'abord, et ensuite pour chaque classe de trafic. Nous obtenons les prix optimaux pour chaque station :

$$u_1^* = 0.1623, \quad u_2^* = 0.1623, \quad u_3^* = 0.5125 \quad \text{et} \quad u_4^* = 0.0638,$$

les prix optimaux pour chaque type de trafic :

$$u^{1*} = 0.5467 \quad \text{et} \quad u^{2*} = 0.5955.$$



## 4.6 Conclusion

Nous avons analysé un modèle de tarification dans un réseau multi-services dont les prix optimisent le surplus collectif du système. Nous avons construit un mécanisme de tarification optimal dans le sens où les utilisateurs vont choisir leur débit de transmission qui maximisent le surplus collectif du système. De plus, notre mécanisme partage le coût de congestion global du système entre les classes de service, ce qui n'est pas vérifié avec une tarification au coût marginal proposé dans la littérature pour ce modèle de réseau à différenciation de service avec une politique d'ordonnancement à priorités strictes.

Une fois notre mécanisme construit, nous nous sommes intéressés à la propriété de compatibilité d'incitation. Cette propriété montre que le mécanisme de tarification incite les utilisateurs à choisir leur classe de service appropriée dans un contexte de classes ouvertes. Nous avons montré que notre mécanisme vérifiait cette propriété pour des temps de service homogènes. Puis, nous l'avons modifié en prenant en compte les temps de service afin que la propriété de compatibilité d'incitation soit vérifiée dans le cas de temps de service non-homogènes.

Nous avons ensuite proposé des extensions de notre modèle en suivant les extensions proposées dans la littérature pour le mécanisme de tarification au coût marginal. On s'est intéressé tout d'abord à un mécanisme de tarification dynamique, où les débits varient sur des périodes de temps en fonction du passé. Enfin, nous avons étendu notre modèle au cas d'un réseau ouvert de Jackson.



## Chapitre 5

# Tarification avec mesures actives et passives des performances

Dans le chapitre précédent, les utilisateurs sont supposés adapter leur débit au taux optimal  $\lambda^*$ . Mais, les taux optimaux ne correspondent pas forcément à la consommation réelle des utilisateurs. Or, un mécanisme de tarification à l'usage nécessite une connaissance exacte de la consommation.

Nous supposons dans ce chapitre que cette quantité est mesurée par le fournisseur. On utilise pour cela des techniques de métrologie des réseaux. Ces mécanismes ont pour objectif d'étudier et de caractériser le trafic existant sur les liens d'Internet en se basant sur des mesures de trafic. On différencie en deux familles les mécanismes de mesure : les mesures *passives* et les mesures *actives*.

Le principe de la métrologie passive est d'observer le trafic sans le perturber. Il existe des logiciels simples pour pouvoir réaliser ces études comme TSTAT [80]. Les statistiques collectées concernent : le protocole utilisé par le paquet, le débit, le temps de passage ou encore les interarrivées. La métrologie active (ou intrusive) [90] permet d'obtenir des mesures du trafic par l'envoi de sonde à travers le lien étudié et recueillir les statistiques sur ces paquets. Elle permet de connaître des mesures de performances telles le délai, la probabilité de perte, la gigue ou encore la route. L'inconvénient de cette approche est la perturbation introduite par les sondes qui augmentent l'intensité du trafic et également les problèmes de synchronisation d'horloges [83]. Un exemple simple d'outil logiciel qui utilise cette approche est le *ping* qui permet de valider un chemin entre deux nœuds d'un réseau et rend quelques statistiques comme le temps de transmission aller-retour et la probabilité de perte.

Dans ce chapitre, le système utilise un mécanisme de mesures passives afin d'estimer la consommation des classes de trafic. Comme dans le chapitre précédent, le système définit les prix optimaux afin de forcer les utilisateurs à choisir leurs débits qui maximisent le surplus collectif. Ces prix optimaux dépendent explicitement du délai moyen des paquets. Ce dernier doit donc être également estimé par le système et pour cela nous supposons que le fournisseur utilise des mesures actives comme celles spécifiées dans [19].

Pour simplifier l'analyse, nous proposons d'étudier dans ce chapitre une tarification au coût marginal, mais on pourrait également utiliser un mécanisme comme proposé dans le chapitre précédent.

## 5.1 Modélisation

Les utilisateurs sont supposés sensibles à leur prix, à leur délai de transmission moyen mais également à l'erreur commise par le système pour l'estimation de leur consommation. On introduit alors une (dés)-utilité pour mauvaises mesures. Le système doit déterminer les paramètres de mesures adéquates pour avoir de bonnes estimations à un moindre coût. En effet, les mesures passives possèdent un coût non-négligeable (traitement, stockage,...) que nous supposons linéaire en fonction de la quantité de trafic observée par le système. Cette proportion du trafic global doit être suffisamment importante pour avoir une bonne estimation de la consommation des utilisateurs car plus l'échantillon est important, plus l'estimateur est précis (faible variance). Il existe donc un compromis sur la quantité de trafic à échantillonner. Quant aux mesures actives, le système doit répondre également à un compromis. En effet, le taux d'envoi des sondes doit être assez important pour obtenir une estimation correcte du délai moyen mais l'intrusion d'un trafic trop important va congestionner le système, entraînant l'augmentation du délai moyen et donc du coût des utilisateurs.

Le système étudié, un goulot d'étranglement du réseau par exemple, est modélisé par une file d'attente M/M/1/PAPS comme dans [6] et dans les chapitres précédents. Nous supposons que tous les paquets sont de taille identique pour les  $K$  classes de trafic ou utilisateurs. Le processus d'arrivée des paquets de classe  $i \in \{1, \dots, K\}$  est un processus de Poisson de taux  $\lambda_i$  et le taux de service est  $\mu$ . On note  $\mathbb{E}D$  le délai moyen des paquets dans la file. Chaque utilisateur de classe  $i$  est caractérisé par la fonction de valuation marginale  $V'_i(\lambda_i)$  qui dépend du taux de transmission  $\lambda_i$ . Un utilisateur de classe  $i$  perçoit un coût d'utilisation du système composé

- du prix par paquet  $u_i$ ,
- du coût linéaire en fonction du délai moyen  $v_i \mathbb{E}D$ ,
- et du coût de l'erreur sur l'estimation du prix  $\beta_i \sigma_i$ . Celui-ci est supposé proportionnel à la variance de l'estimateur du prix, donc égal à  $\beta_i \sigma_i$ , avec  $\sigma_i = \sqrt{\mathbb{V}(\hat{u}_i)}$ . En effet, le prix dépend de la consommation des utilisateurs laquelle est mesurée par le système. Donc, l'erreur sur la mesure de consommation induit une erreur sur le prix.

Le coût total perçu par un utilisateur de classe  $i$  est donc :  $u_i + v_i \mathbb{E}D + \beta_i \sigma_i$ . Un utilisateur choisira de soumettre son trafic et ainsi augmente son débit, si et seulement si sa valuation excède son coût total. Ainsi, à l'équilibre, nous avons la relation de demande suivante :

$$V'_i(\lambda_i) = u_i + v_i \mathbb{E}D + \beta_i \sigma_i. \quad (5.1)$$

Cette relation montre comment le débit d'une classe d'application s'adapte en fonction des coûts et permet alors de connaître le comportement des utilisateurs. Nous notons

$\lambda = \sum_{i=1}^K \lambda_i$ , le taux d'arrivée total dans le système et  $\underline{\lambda} = (\lambda_1, \dots, \lambda_K)$  le vecteur des taux d'arrivée par classe. Le système est capable à partir de la relation de demande (5.1) de déterminer les prix optimaux qui forcent les utilisateurs à choisir les débits qui optimisent le surplus collectif du système.

## 5.2 Tarification Optimale

L'objectif du système est de déterminer les prix et les paramètres de mesure qui optimisent le surplus collectif. Nous supposons que le système réalise des mesures passives en échantillonnant une proportion  $\varepsilon$  du trafic total. On suppose un coût  $\zeta$  par unité de trafic échantillonné. Le système possède un coût  $\zeta\varepsilon\lambda$  supposé linéaire en fonction de la taille de l'échantillon. Le surplus collectif pour le système est alors :

$$V(\lambda) - L(\lambda) - \zeta\varepsilon\lambda. \quad (5.2)$$

Le théorème suivant détermine le vecteur  $\underline{u}^*$  des prix qui optimisent (indirectement via les débits) le surplus collectif du système.

**Théorème 13** *Le prix  $u_i^*$  pour le trafic de classe  $i$  qui maximise le surplus collectif est donné par*

$$\forall i \in \{1, \dots, K\}, \quad u_i^* = \mathbb{E}D^2 \sum_{k=1}^K v_k \lambda_k^* - \beta_i \sigma_i + \zeta\varepsilon,$$

avec  $\underline{\lambda} = \underline{\lambda}^*$  le taux d'arrivé qui maximise l'équation (5.2).

**Preuve** Le vecteur des taux d'arrivées  $\underline{\lambda}^*$  qui optimise la valuation totale du réseau (5.2) vérifie :

$$\forall i = 1, \dots, K, \quad V_i'(\lambda_i^*) = v_i \mathbb{E}D(\underline{\lambda}^*) + \sum_{k=1}^K v_k \lambda_k^* \frac{\partial \mathbb{E}D}{\partial \lambda_i}(\underline{\lambda}^*) + \zeta\varepsilon.$$

De plus, nous avons la relation de demande décrite par l'équation 5.1 qui est :

$$\forall i = 1, \dots, K, \quad V_i'(\lambda_i^*) = u_i + v_i \mathbb{E}D + \beta_i \sigma_i.$$

En combinant ces deux équations, nous obtenons le vecteur  $\underline{u}^*$  des prix en fonction des taux d'arrivées optimaux

$$\forall i = 1, \dots, K, \quad u_i^* = \sum_{k=1}^K v_k \lambda_k^* \frac{\partial \mathbb{E}D}{\partial \lambda_i}(\underline{\lambda}^*) - \beta_i \sigma_i + \zeta\varepsilon. \quad (5.3)$$

Le système est modélisé par une file d'attente M/M/1/PAPS. Ainsi, le délai moyen est vérifiée pour tout  $i$ ,  $\frac{\partial \mathbb{E}D}{\partial \lambda_i} = \mathbb{E}D^2$ . Donc, pour tout  $i = 1, \dots, K$ , le prix optimal  $u_i^*$  pour la  $i^{ieme}$  classe de service est d'après (5.3)

$$u_i^* = \mathbb{E}D^2 \sum_{k=1}^K v_k \lambda_k^* - \beta_i \sigma_i + \zeta\varepsilon. \quad (5.4)$$

■

L'expression  $v_k \lambda_k^* \frac{\partial \mathbb{E}D}{\partial \lambda_i} = v_k \lambda_k^* \mathbb{E}D^2$  représente le coût d'externalité d'un utilisateur de classe  $i$  sur un utilisateur de classe  $k$ . Donc, le prix optimal  $u_i^*$  se compose de la somme des externalités induites par le trafic de classe  $i$  sur les autres classes (qui est le prix marginal du modèle de Mendelson et Whang [81]) et du coût marginal  $\zeta \epsilon$  des mesures passives effectuées par le système. On remarque que le coût  $\beta_i \sigma_i$  de l'erreur d'estimation n'apparaît pas pour le mécanisme du prix marginal. En effet, cette erreur est due aux mesures effectuées par le système et n'est pas facturée.

Les prix optimaux  $\underline{u}^*$  dépendent des débits optimaux  $\underline{\lambda}^*$  et du délai moyen  $\mathbb{E}D$ . Ces deux mesures de performance du système sont évaluées par le système via des techniques liées à la métrologie des réseaux de communication.

### 5.3 Mesures actives et passives des performances

Les deux paramètres de qualité de service qui nous intéressent sont le débit et le délai moyen de bout en bout. Dans les réseaux de communication, le délai moyen est généralement estimé par des techniques de mesure active, c'est-à-dire en injectant un trafic spécifique dans le système [38]. Ce trafic est traité de la même façon que le trafic réel mais sert uniquement à mesurer le temps de réponse. Ainsi, le délai moyen obtenu sur les sondes estime le délai moyen du système. Il existe plusieurs façon de générer ce trafic. Dans [20] par exemple, les sondes sont générées périodiquement et ainsi, en faisant varier l'intervalle de temps entre les sondes, l'auteur étudie la structure de l'Internet. D'une autre façon pour notre modèle, nous supposons que les sondes sont générées suivant un processus de Poisson de taux  $\gamma$ , afin de conserver des arrivées poissonniennes et donc une file M/M/1. Nous avons donc une boîte noire en entrée et en sortie du système qui envoie des sondes suivant un processus de Poisson et les récupère en sortie. Le délai moyen est estimé en faisant la moyenne des temps de traitement de ces sondes par le système. On les suppose de taille égale aux paquets générés par les utilisateurs. Quant au débit moyen  $\lambda_i$  d'arrivée des paquets de classe  $i$ , le système observe le trafic réel (sans les sondes) entrant dans le système. En raison de coûts importants liés au stockage et à l'analyse des paquets, nous supposons que seule une proportion  $\epsilon$  du trafic peut être observée. Ceci est régulièrement supposé dans les techniques d'estimation du débit avec des mesures passives comme par exemple dans [68, 90]. Chaque paquet joignant le système est échantillonné suivant une loi de Bernoulli de moyenne  $\epsilon$  (indépendante entre les paquets). De plus, nous supposons que les paquets de mesures actives sont automatiquement reconnus par le système et ne peuvent être échantillonnés (voir Figure 5.1).

Le temps est décomposé en intervalles fixes de  $T$  unités de temps. Ces périodes sont appelées périodes de mesure. Le nombre de classe d'utilisateurs  $K$  est supposé fixe durant ces périodes tout comme leur taux d'arrivée  $\lambda_i$  pour  $i \in \{1, \dots, K\}$ . Les estimateurs sont calculés en fonction des mesures effectuées tout au long de la période. On dénote par  $N$  le nombre total de paquets de trafic réel (sans les sondes) joignant le système durant une période de mesure. On pose également  $N_s$  le nombre de paquets échantillonnés durant une période et les variables de Bernoulli  $X_{ij}$  avec  $1 \leq j \leq N_s$  et

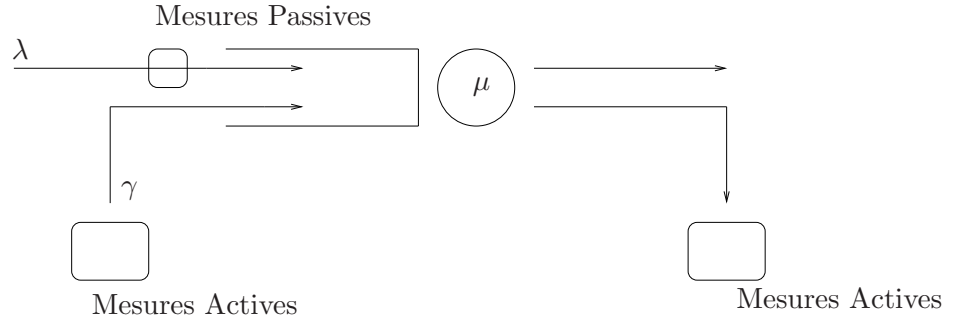


FIG. 5.1 – Modèle avec outils de mesures actives et passives.

$1 \leq i \leq K$  sont définies par :

$$X_{ij} = \begin{cases} 1 & \text{si le } j^{\text{ème}} \text{ paquet échantillonné appartient à la classe } i, \\ 0 & \text{sinon.} \end{cases}$$

A partir de ces variables, nous construisons un estimateur sans-biais du débit de transmission  $\lambda_i$  pour chaque classe  $i$ .

### 5.3.1 Estimation du débit

Dans la proposition suivante nous construisons l'estimateur (passif)  $\hat{\lambda}_i$  du débit des paquets de classe  $i$  avec  $i \in \{1, \dots, K\}$ .

**Proposition 7** *On définit l'estimateur suivant :*

$$\hat{\lambda}_i = \frac{\sum_{j=1}^{N_s} X_{ij}}{\varepsilon T}. \quad (5.5)$$

*Cet estimateur est sans-biais et sa variance est*

$$\mathbb{V}[\hat{\lambda}_i] = \frac{\lambda_i}{\varepsilon T}.$$

**Preuve** Le nombre de paquets  $N$  joignant le système durant une période de mesure de longueur  $T$  suit une distribution de Poisson de taux  $\lambda T$ . En conditionnant à  $N$ , la variable aléatoire  $N_s$  suit une distribution binomiale  $\mathcal{B}(N, \varepsilon)$ . Observons la fonction génératrice de la variable  $\varepsilon T \hat{\lambda}_i = \sum_{i=1}^{N_s} X_{ij}$  :

$$\begin{aligned} G_{\varepsilon T \hat{\lambda}_i}(s) &= \mathbb{E}[\mathbb{E}[s^{\sum_{k=1}^{N_s} X_{ki}} | N_s]] &= \mathbb{E}[(1 - \rho_i) + s\rho_i]^{N_s} \\ &= \mathbb{E}[\mathbb{E}[(1 - \rho_i) + s\rho_i]^{N_s} | N]] \\ &= \mathbb{E}[(1 - \varepsilon) + ((1 - \rho_i) + s\rho_i)\varepsilon]^N \\ &= \exp(-\lambda T((1 - \varepsilon) + ((1 - \rho_i) + s\rho_i)\varepsilon - 1)) \\ &= \exp(\lambda_i T \varepsilon (s - 1)). \end{aligned}$$

La variable aléatoire  $\varepsilon T \hat{\lambda}_i$  suit donc une loi de Poisson de taux  $\lambda_i T \varepsilon$ , d'où :

$$\mathbb{E}(\varepsilon T \hat{\lambda}_i) = \varepsilon T \mathbb{E}(\hat{\lambda}_i) = \lambda_i T \varepsilon.$$

L'estimateur  $\hat{\lambda}_i$  est donc un estimateur sans-biais du taux d'envoi  $\lambda_i$ . Sa variance est définie par :

$$\mathbb{V}[\hat{\lambda}_i] = \frac{1}{\varepsilon^2 T^2} \mathbb{V}[\varepsilon T \hat{\lambda}_i] = \frac{\lambda_i T \varepsilon}{\varepsilon^2 T^2} = \frac{\lambda_i}{\varepsilon T}.$$

■

### 5.3.2 Estimation du délai moyen

Le délai moyen est estimé en injectant un trafic dans le système suivant un processus de Poisson de taux  $\gamma$ . Ainsi, le taux d'arrivée global dans la file est  $\lambda + \gamma$ . Le temps de réponse pour une file d'attente M/M/1 suit une variable aléatoire exponentielle [63] dont le paramètre est la différence entre le taux de service et le taux d'arrivée, si cette différence positive. Pour notre modèle, le délai suit donc une loi exponentielle de taux  $\mu - \lambda - \gamma$  (il est requis que  $\gamma < \mu - \lambda$ ). De plus, nous supposons que le taux d'envoi des sondes  $\gamma$  est très inférieur à  $\lambda$  et ainsi, comme proposé dans [90], les délais des sondes sont approximativement indépendants. Nous pouvons maintenant construire notre estimateur de  $\mathbb{E}D$ . Pour cela, on note  $N_a$  le nombre de sondes envoyées dans le système durant une période de mesure.

On définit l'estimateur du délai moyen suivant :

$$\hat{d} = \frac{\sum_{k=1}^{N_a} d_k}{\gamma T}, \quad (5.6)$$

si  $N_a > 0$  et 0 sinon. La variable aléatoire  $d_k$  représente le temps de séjour de la  $k^{eme}$  sonde dans le système. Le nombre  $N_a$  de sondes envoyées pendant un intervalle de temps de durée  $T$  suit une distribution de Poisson de paramètre  $\gamma T$ . Le moment d'ordre  $r$  de cet estimateur est donné par l'expression suivante :

$$\begin{aligned} \mathbb{E}((\gamma T \hat{d})^r) &= 0 \mathbb{P}(N_a = 0) + \mathbb{E} \left( \left( \sum_{k=1}^{N_a} d_k \right)^r \right), \\ &= \sum_{n_a=0}^{\infty} \mathbb{E} \left( \left( \sum_{k=1}^{n_a} d_k \right)^r \right) \mathbb{P}(N_a = n_a). \end{aligned}$$

L'idée est que pour un nombre de sonde assez petit par rapport au nombre de paquet de trafic des utilisateurs et pour une durée  $T$  des périodes de mesure élevée, on peut supposer l'indépendance des temps de séjour des sondes et une approximation de la loi de leur temps de séjour. Ainsi nous obtenons une valeur approchée simple du moment d'ordre  $r$  de cet estimateur. Pour un nombre de sondes plus élevé pendant la



période de temps  $T$ , on va borner la différence entre le moment d'ordre  $r$  de l'estimateur et l'approximation. Pour cela, on se fixe l'entier  $m$  qui correspond à un nombre petit de sonde pour lesquels l'indépendance des sondes peut être approchée et également on pourra supposer que le temps de séjour approche une loi exponentielle. On se fixe également un réel  $\delta$  (petit) qui va majorer le reste de la différence entre le moment d'ordre  $r$  de l'estimateur et son approximation. On détermine alors l'ensemble des paramètres  $\gamma$  et  $T$  tel que

$$\sum_{n_a=m+1}^{\infty} (n_a T)^r \frac{(\gamma T)^{n_a}}{n_a!} e^{-\gamma T} \leq \frac{\delta}{2}. \quad (5.7)$$

Comme cette série converge, il faut que le produit  $\gamma T$  soit assez grand pour que l'exponentielle écrase la queue de cette série et qu'elle puisse être alors majorée par  $\delta/2$ . On suppose de plus que la durée  $T$  vérifie :

$$T > \max \left( \frac{1}{\gamma}, \frac{1}{\mu - \lambda - \gamma} \sqrt[r]{\frac{(m+r) \dots (m+1)}{m^r}} \right). \quad (5.8)$$

Nous avons donc

$$T^r > \frac{1}{(\mu - \lambda - \gamma)^r} \left( 1 + \frac{a_1}{m} + \dots + \frac{a_r}{m^r} \right),$$

avec  $a_1, \dots, a_r$  des constantes. Or pour tout  $n_a > m$  nous avons :

$$\forall i = 1, \dots, r \quad \frac{1}{m^i} > \frac{1}{n_a^i},$$

et donc :

$$T^r > \frac{1}{(\mu - \lambda - \gamma)^r} \left( 1 + \frac{a_1}{n_a} + \dots + \frac{a_r}{n_a^r} \right).$$

Alors pour tout  $n_a > m$ , on a :

$$(n_a T)^r > \frac{1}{(\mu - \lambda - \gamma)^r} (n_a + r) \dots (n_a + 1). \quad (5.9)$$

A partir de la majoration (5.7) et de la relation (5.9), nous avons la majoration :

$$\sum_{n_a=m+1}^{\infty} \frac{(n_a + r) \dots (n_a + 1)}{(\mu - \lambda - \gamma)^r} \frac{(\gamma T)^{n_a}}{n_a!} e^{-\gamma T} \leq \frac{\delta}{2}. \quad (5.10)$$

On sépare alors le calcul du moment d'ordre  $r$  de l'estimateur en fonction de l'entier  $m$  défini auparavant. On obtient alors :

$$\mathbb{E}((\gamma T \hat{d})^r) = \sum_{n_a=1}^m \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) \mathbb{P}(N_a = n_a) + \sum_{n_a=m+1}^{\infty} \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) \mathbb{P}(N_a = n_a).$$

Comme le taux d'arrivée des sondes est faible par rapport au taux d'arrivée des utilisateurs, pendant une durée  $T$  il y aura beaucoup de trafic d'utilisateurs ou même personne dans la file entre deux sondes successives. Ainsi, on peut supposer l'indépendance

des temps de séjour des sondes. De plus, comme  $T$  est supposé grand par rapport au temps moyen d'interarrivées entre deux sondes successives, la distribution du temps de séjour approche une loi exponentielle.

Avec ces hypothèses d'indépendance et de distribution exponentielle, nous avons  $\sum_{k=1}^{n_a} d_k$  approche une loi Gamma de paramètre  $(n_a, \mu - \lambda - \gamma)$ . Le moment d'ordre  $r \geq 1$  tronqué à  $n_a$  sondes avec  $n_a \in \{1, \dots, m\}$  approche :

$$\mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) \simeq \frac{\Gamma(n_a + r)}{(\mu - \lambda - \gamma)^r \Gamma(n_a)}.$$

Or  $\Gamma(n_a + r) = (n_a + r - 1)!$  pour tout  $r \geq 1$ . On en déduit alors une approximation du moment d'ordre  $r$  de  $\gamma T \hat{d}$  :

$$\begin{aligned} \mathbb{E}\left((\gamma T \hat{d})^r\right) &\simeq \sum_{n_a=1}^m \frac{(n_a + r)!}{(\mu - \lambda - \gamma)^r n_a!} \mathbb{P}(N_a = n_a) + \sum_{n_a=m+1}^{\infty} \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) \mathbb{P}(N_a = n_a), \\ &= \frac{1}{(\mu - \lambda - \gamma)^r} \left( \mathbb{E}\left(\frac{(N_a + r)!}{N_a!}\right) - \sum_{n_a=m+1}^{\infty} \frac{(n_a + r)!}{n_a!} \mathbb{P}(N_a = n_a) \right) \\ &\quad + \sum_{n_a=m+1}^{\infty} \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) \mathbb{P}(N_a = n_a), \\ &= \frac{\mathbb{E}\left(\frac{(N_a + r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r} + \sum_{n_a=m+1}^{\infty} \left( \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) - \frac{(n_a + r)!}{(\mu - \lambda - \gamma)^r n_a!} \right) \mathbb{P}(N_a = n_a). \end{aligned}$$

On obtient alors :

$$\begin{aligned} \left| \mathbb{E}\left((\gamma T \hat{d})^r\right) - \frac{\mathbb{E}\left(\frac{(N_a + r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r} \right| &\simeq \left| \sum_{n_a=m+1}^{\infty} \left( \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) - \frac{(n_a + r)!}{(\mu - \lambda - \gamma)^r n_a!} \right) \mathbb{P}(N_a = n_a) \right|, \\ &< \sum_{n_a=m+1}^{\infty} \left| \mathbb{E}\left(\left(\sum_{k=1}^{n_a} d_k\right)^r\right) - \frac{(n_a + r)!}{(\mu - \lambda - \gamma)^r n_a!} \right| \mathbb{P}(N_a = n_a), \\ &< \sum_{n_a=m+1}^{\infty} \left( (n_a T)^r + \frac{(n_a + r)!}{(\mu - \lambda - \gamma)^r n_a!} \right) \frac{(\gamma T)^{n_a}}{n_a!} e^{-\gamma T}, \\ &< \sum_{n_a=m+1}^{\infty} (n_a T)^r \frac{(\gamma T)^{n_a}}{n_a!} e^{-\gamma T} + \sum_{n_a=m+1}^{\infty} \frac{(n_a + r)!}{(\mu - \lambda - \gamma)^r n_a!} \frac{(\gamma T)^{n_a}}{n_a!} e^{-\gamma T}. \end{aligned}$$

D'après les majorations (5.7) et (5.10), la différence absolue entre le moment d'ordre  $r$  de l'estimateur du délai et de son approximation est bornée par :

$$\begin{aligned} \left| \mathbb{E}\left((\gamma T \hat{d})^r\right) - \frac{\mathbb{E}\left(\frac{(N_a + r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r} \right| &< \frac{\delta}{2} + \frac{\delta}{2}, \\ &< \delta. \end{aligned}$$

Nous avons donc

$$\left| \mathbb{E}((\gamma T \hat{d})^r) - \frac{\mathbb{E}\left(\frac{(N_a+r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r} \right| < \frac{\delta}{(\gamma T)^r}.$$

Or d'après (5.8), nous avons  $T > 1/\gamma$  et donc :

$$\left| \mathbb{E}((\gamma T \hat{d})^r) - \frac{\mathbb{E}\left(\frac{(N_a+r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r} \right| < \delta.$$

Ce résultat est vrai pour tout  $\delta$  et donc on peut approcher le moment d'ordre  $r$  de l'estimateur du temps de réponse par  $\frac{\mathbb{E}\left(\frac{(N_a+r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r}$ , i.e.

$$\mathbb{E}((\gamma T \hat{d})^r) \simeq \frac{\mathbb{E}\left(\frac{(N_a+r)!}{N_a!}\right)}{(\mu - \lambda - \gamma)^r}. \quad (5.11)$$

Alors la variance de cet estimateur est approchée par :

$$\begin{aligned} \mathbb{V}(\hat{d}) &\simeq \mathbb{E}(\hat{d}^2) - [\mathbb{E}(\hat{d})]^2, \\ &\simeq \frac{\gamma T + 2}{\gamma T (\mu - \lambda - \gamma)^2} - \frac{1}{(\mu - \lambda - \gamma)^2}, \\ &\simeq \frac{2}{\gamma T (\mu - \lambda - \gamma)^2}. \end{aligned}$$

Nous avons ainsi construit un estimateur sans-biais  $\hat{\lambda}_i$  du débit pour chaque classe  $i$  et un estimateur  $\hat{d}$  qui approche le délai moyen perçu par les paquets dans le système. Ces estimateurs sont basés sur deux différentes techniques de mesures : *actives* pour estimer le délai et *passives* pour le débit.

Le système règle le taux d'envoi des mesures actives très inférieur au débit des utilisateurs, c'est-à-dire  $\gamma \ll \lambda$  afin que les délais de traitement des mesures soient approximativement indépendants.

Nous pouvons donc maintenant déterminer en fonction des paramètres de mesure  $\varepsilon$  et  $\gamma$ , la variance  $\sigma_i$  de l'estimateur du prix qui intervient dans le calcul du prix optimal  $u_i^*$  défini par le théorème (13).

**Proposition 8** *La variance de l'estimateur  $\hat{u}_i$  du prix pour la  $i^{eme}$  classe est*

$$\sigma_i^2 = \frac{\hat{d}^4}{T} \left[ \left( 1 + \frac{12}{\gamma T} + \frac{36}{\gamma^2 T^2} + \frac{24}{\gamma^3 T^3} \right) \sum_{i=1}^K v_i^2 \frac{\hat{\lambda}_i}{\varepsilon} + \frac{8}{\gamma} \left( 1 + \frac{4}{\gamma T} + \frac{3}{\gamma^2 T^2} \right) \left( \sum_{i=1}^K v_i \hat{\lambda}_i \right)^2 \right].$$

**Preuve** L'erreur d'estimation du prix est définie par :

$$\sigma_i^2 = \text{Var}(\hat{u}_i) = \text{Var}\left(\hat{d}^2 \sum_{k=1}^K v_k \hat{\lambda}_k - \beta_i \sigma_i + \zeta \epsilon\right) = \text{Var}\left(\hat{d}^2 \sum_{k=1}^K v_k \hat{\lambda}_k\right).$$

Les sondes ne sont pas pris en compte dans le trafic observé pour l'estimation du débit (voir figure 5.1) et les estimateurs  $\hat{d}$  et  $\hat{\lambda}_k$  sont donc indépendants pour tout  $k \in \{1, \dots, K\}$ . Donc l'erreur  $\sigma_i$  est

$$\sigma_i = \mathbb{E}(\hat{d}^4) \mathbb{V}\left(\sum_{k=1}^K v_k \hat{\lambda}_k\right) + \mathbb{E}\left(\sum_{k=1}^K v_k \hat{\lambda}_k\right)^2 \mathbb{V}(\hat{d}^2).$$

Or nous avons d'après (5.11) :

$$\begin{aligned} \mathbb{E}(\hat{d}^4) &= \frac{1}{\gamma^3 T^3 (\mu - \lambda - \gamma)^4} (\gamma^3 T^3 + 12\gamma^2 T^2 + 36\gamma T + 24), \\ \mathbb{V}(\hat{d}^2) &= \frac{8}{\gamma T (\mu - \lambda - \gamma)^4} \left(1 + \frac{4}{\gamma T} + \frac{3}{\gamma^2 T^2}\right), \end{aligned}$$

et de plus

$$\begin{aligned} \mathbb{V}\left(\sum_{k=1}^K v_k \hat{\lambda}_k\right) &= \sum_{k=1}^K v_k^2 \frac{\lambda_k}{\epsilon T}, \\ \mathbb{E}\left(\sum_{k=1}^K v_k \hat{\lambda}_k\right) &= \sum_{k=1}^K v_k \lambda_k. \end{aligned}$$

Ceci implique après simplifications, que l'erreur sur l'estimation du prix optimal est égale à :

$$\sigma_i^2 = \frac{\mathbb{E}D^4}{T} \left[ \left(1 + \frac{12}{\gamma T} + \frac{36}{\gamma^2 T^2} + \frac{24}{\gamma^3 T^3}\right) \sum_{i=1}^K v_i^2 \frac{\lambda_i^*}{\epsilon} + \frac{8}{\gamma} \left(1 + \frac{4}{\gamma T} + \frac{3}{\gamma^2 T^2}\right) \left(\sum_{i=1}^K v_i \lambda_i^*\right)^2 \right].$$

Donc, nous considérons l'estimateur biaisé de cette variance, sans distinction de notations, par :

$$\sigma_i^2 = \frac{\hat{d}^4}{T} \left[ \left(1 + \frac{12}{\gamma T} + \frac{36}{\gamma^2 T^2} + \frac{24}{\gamma^3 T^3}\right) \sum_{i=1}^K v_i^2 \frac{\hat{\lambda}_i}{\epsilon} + \frac{8}{\gamma} \left(1 + \frac{4}{\gamma T} + \frac{3}{\gamma^2 T^2}\right) \left(\sum_{i=1}^K v_i \hat{\lambda}_i\right)^2 \right]. \quad \blacksquare$$

On remarque que cette erreur est décroissante avec  $T$ . En effet, plus le système réalise des mesures, plus elles sont précises. Les prix optimaux déterminés dans le théorème 13 implique la maximisation du surplus collectif du système. En effet, les prix  $\underline{u}^* = (u_1^*, \dots, u_K^*)$  forcent les taux d'arrivées  $\lambda_i^*$  par la relation de demande (5.1). Le système doit également déterminer les paramètres de mesures optimaux  $\epsilon^*$  et  $\gamma^*$  qui optimisent le surplus collectif. Dans la section suivante nous donnons plusieurs versions d'un même algorithme d'ajustement des paramètres optimaux de mesures pour chaque période de mesure. Les algorithmes diffèrent par la méthode de prise en compte des mesures du passé dans la future estimation.

## 5.4 Procédure d'estimation des performances

L'évolution du système se fait par slot de temps. Les estimations effectuées par le fournisseur induisent un changement dans le comportement des utilisateurs pour la prochaine période de temps. Nous décrivons sur une période de mesure  $t$ , les mécanismes mis en œuvre d'un point de vue séquentiel. Pour cela, nous nous plaçons à la fin de la  $t^{ieme}$  période de mesure avec des paramètres de mesures  $\epsilon_t$  et  $\gamma_t$  connus et décrivons la procédure d'estimation des performances du réseau.

1. A partir des estimateurs (5.5) et (5.6), le système détermine  $\hat{\lambda}_i^t, \forall i = 1, \dots, K$ , et  $\hat{d}^t$  les estimations du débit pour chaque classe et du délai moyen pour la  $t^{ieme}$  période de mesure.
2. Pour  $t > 1$ , les estimations du passé, pour les périodes de 1 à  $t - 1$  sont combinés avec l'estimateur courant de sorte que  $\forall i \in \{1, \dots, I\}$ ,

$$\tilde{\lambda}_i^t = r_\lambda^t \frac{\sum_{k=1}^{t-1} \hat{\lambda}_i^k}{t-1} + (1 - r_\lambda^t) \hat{\lambda}_i^t,$$

et,

$$\tilde{d}^t = r_d^t \frac{\sum_{k=1}^{t-1} \hat{d}^k}{t-1} + (1 - r_d^t) \hat{d}^t,$$

avec  $r_\lambda^t$  et  $r_d^t$  sont des réels positifs plus petits que 1. Ces derniers représentent le poids dans le nouvel estimateur associé au passé. On note que ces estimateurs sont des combinaisons linéaires d'estimateurs sans-biais et sont donc également non biaisés.

3. Les prix pour la prochaine période  $t + 1$  sont calculés avec l'expression donnée dans le théorème 13, c'est-à-dire

$$\forall i \in \{1, \dots, K\} \quad \hat{u}_i^{t+1} = (\tilde{d}^t)^2 \sum_{i=1}^K v_i \tilde{\lambda}_i^t - \beta_i \sigma_i(\tilde{\lambda}^t, \tilde{d}^t, \gamma_t, \epsilon_t) + \zeta \epsilon_t,$$

avec

$$\begin{aligned} \sigma_i^2(\underline{\lambda}, d, \gamma, \epsilon) &= \frac{d^4}{T} \left( \left( 1 + \frac{12}{\gamma T} + \frac{36}{\gamma^2 T^2} + \frac{24}{\gamma^3 T^3} \right) \sum_{i=1}^K v_i^2 \frac{\lambda_i}{\epsilon} \right. \\ &\quad \left. + \frac{8}{\gamma} \left( 1 + \frac{4}{\gamma T} + \frac{3}{\gamma^2 T^2} \right) \left( \sum_{i=1}^K v_i \lambda_i \right)^2 \right). \end{aligned}$$

4. Le système détermine les paramètres de mesures  $\epsilon_{t+1}$  et  $\gamma_{t+1}$  pour la période suivante qui maximisent le surplus collectif du système, c'est-à-dire :

$$(\gamma_{t+1}, \epsilon_{t+1}) = \arg \max_{\gamma, \epsilon} \sum_{i=1}^K \left( V_i(\lambda_i^{t+1}) - \frac{v_i \lambda_i^{t+1}}{\mu - \sum_{i=1}^K \lambda_i^{t+1} - \gamma} - \zeta \epsilon \lambda_i^{t+1} \right),$$

en supposant que la future demande (débit) vérifie la relation de demande (5.1) :

$$V'_i(\lambda_i^{t+1}) = \hat{u}_i^{t+1} + v_i \tilde{d}^t + \beta_i \sigma_i(\tilde{\lambda}^t, \tilde{d}^t, \gamma, \epsilon).$$

On approche le délai moyen de la prochaine période  $t + 1$  par le délai en régime stationnaire de la file d'attente.

Dans la dernière étape, le système approche le comportement des utilisateurs de classe  $i$  par la fonction de valuation  $V'_i$ . Nous proposons trois algorithmes qui diffèrent suivant les valeurs associées aux poids  $r_\lambda^t$  et  $r_d^t$ .

## 5.5 Algorithmes d'ajustement des paramètres de mesures

Nous considérons trois algorithmes différents pour la construction des nouveaux estimateurs  $\tilde{\lambda}_i^t$  et  $\tilde{d}^t$  suivant la façon de prendre en compte les estimations effectuées dans le passé. Cette technique de moyenne pondérée est similaire à celle utilisée dans les routeurs internes d'un domaine *DiffServ* pour améliorer l'estimation de la taille moyenne de la file pour les algorithmes de gestion active des files d'attente comme RED [44]. Nous utilisons l'index  $t$  pour indiquer la  $t^{\text{ème}}$  période de mesure.

- *last* : Seule la dernière estimation est prise en compte et donc  $r_d^t = r_\lambda^t = 0 \forall t$ . Ce choix semble idéal dans le cas où la demande varie énormément entre les périodes de mesure et les estimations dans le passé ont donc peu d'intérêt.
- *int* : Pour cet autre algorithme, nous fixons un poids constant non nul pour la dernière estimation, c'est-à-dire  $r_d^t = r_d$  et  $r_\lambda^t = r_\lambda \forall t$ . Ainsi les estimations sont lissées par la prise en compte du passé et de plus, elles sont sensibles à la variation des données à estimer.
- *id* : Pour ce dernier algorithme, le système assure un poids identique pour toutes les estimations, c'est-à-dire  $r_d^t = r_\lambda^t = (t - 1)/t \forall t$ . Cette politique est valable quand les demandes sont constantes dans le temps. En effet, nous avons ainsi :

$$\begin{aligned} \tilde{\lambda}_i^t &= r_\lambda^t \frac{\sum_{k=1}^{t-1} \hat{\lambda}_i^k}{t-1} + (1 - r_\lambda^t) \hat{\lambda}_i^t, \\ &= \frac{t-1}{t} \frac{\sum_{k=1}^{t-1} \hat{\lambda}_i^k}{t-1} + \frac{1}{t} \hat{\lambda}_i^t, \\ &= \frac{\sum_{k=1}^t \hat{\lambda}_i^k}{t}. \end{aligned}$$

Les estimateurs  $\hat{\lambda}_i^k$  sont indépendants et identiquement distribués, on peut donc appliquer la loi forte des grands nombres qui donne :

$$\tilde{\lambda}_i^t = \frac{\sum_{k=1}^t \hat{\lambda}_i^k}{t} \xrightarrow[t \rightarrow +\infty]{} \mathbb{E}(\hat{\lambda}_i^k) = \mathbb{E}(\hat{\lambda}_i) = \lambda_i.$$

Nous obtenons le même résultat de convergence pour l'estimateur du délai moyen.

## 5.6 Applications numériques

Nous illustrons nos algorithmes avec deux types de fonction de valuation. Tout d'abord, nous considérons une fonction de valuation constante dans le temps et nous observons alors que l'algorithme *id* donne les meilleurs résultats. Ensuite, nous supposons que le temps influe sur la valuation et montrons que l'algorithme *int* avec  $r_d = r_\lambda = 0.9$  est préférable aux algorithmes *id* et *last*. Nous considérons  $K = 3$  classes d'utilisateurs et le taux de service  $\mu = 10$ . La durée d'une période de mesure est fixée à  $T = 1000$  unités de temps. Nous considérons que le coût des mesures passives est  $\zeta = 0.005$  et que pour l'initialisation, les paramètres de mesures sont choisis tels que  $\gamma_0$  est égale à 10% de la plus petite demande des classes d'utilisateurs et  $\epsilon_0$  est choisi aléatoirement suivant une distribution uniforme entre 0 et 10% du trafic total.

	Classe 1	Classe 2	Classe 3
$v$	1	1.5	2
$\beta$	2	1	0.5

TAB. 5.1 – Coûts de délai et d'erreurs de mesures pour chacune des classes.

Le tableau 5.1 indique les coûts de délai et d'erreurs de mesures pour chacune des classes. Nous avons choisis ces valeurs afin d'avoir des classes d'utilisateurs hétérogènes. La première classe est la moins sensible au délai mais la plus sensible à l'erreur de mesures et la troisième classe, inversement, est la moins sensible à l'erreur mais par contre est la plus sensible au délai. La deuxième classe se situe entre les deux. Les débits à l'instant initial sont tous arbitrairement égaux à 2.

### 5.6.1 Valuation constante

Les utilisateurs sont supposés posséder une fonction valuation  $V$  constante dans le temps, c'est-à-dire que leur fonction de valuation ne dépend pas de la période de mesure. Nous considérons les fonctions de demande suivantes :

$$\begin{aligned} V_1(\lambda_1) &= 15\lambda_1 - 3\lambda_1^2, \\ V_2(\lambda_2) &= 17\lambda_2 - 4\lambda_2^2, \\ V_3(\lambda_3) &= 16\lambda_3 - 5\lambda_3^2, \end{aligned}$$

définies sur les intervalles respectifs  $[0, 2.5]$ ,  $[0, 2.125]$  et  $[0, 1.6]$ . Nous nous intéressons tout d'abord à l'évolution des débits  $\lambda_i$  ( $i = 1, 2, 3$ ) et au taux d'envoi  $\gamma$  des mesures actives. La figure 5.2(a) montre les résultats pour l'algorithme *last*, la figure 5.2(b) pour l'algorithme *int* avec  $r_d = r_\lambda = 0.9$  et la figure 5.2(c) l'algorithme *id*. La courbe pointillée correspond à l'estimateur et celle en trait plein à la demande réelle.

Lorsque les estimations dans le passé ne sont pas utilisées (algorithme *last*), les estimateurs montrent une grande variabilité car seule la dernière estimation est considérée.

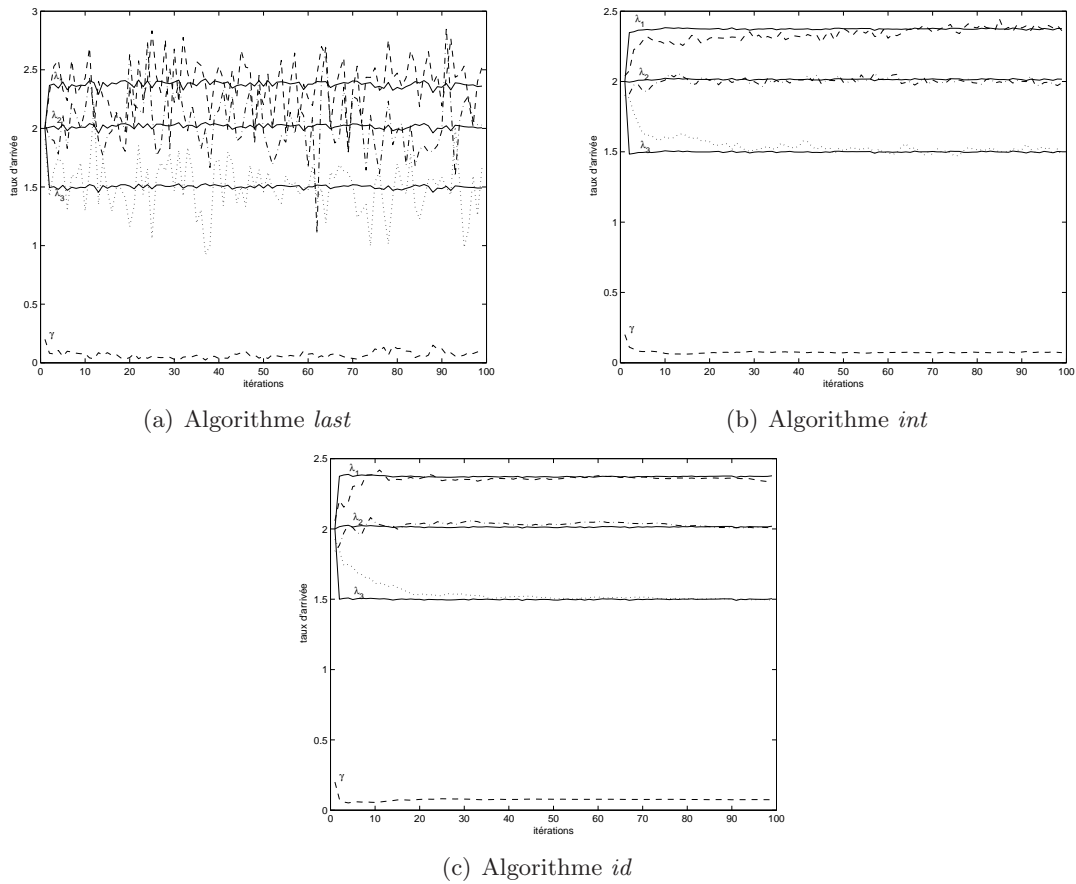
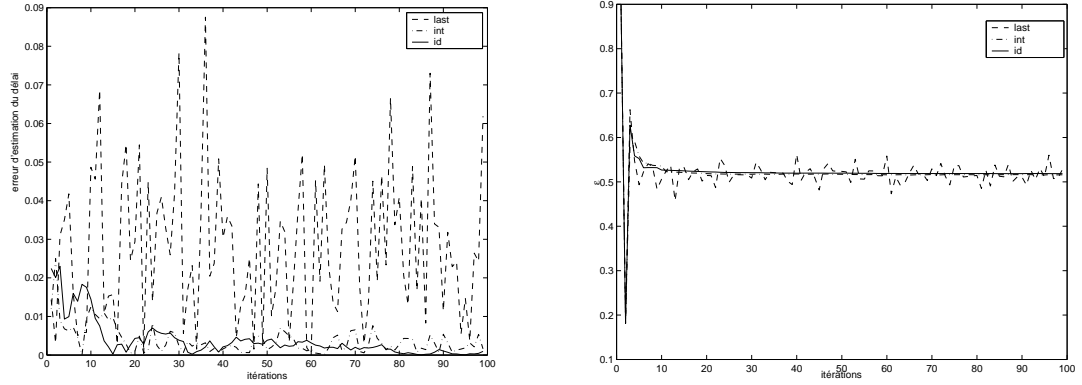


FIG. 5.2 – Évolution des débits  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  et  $\gamma$  au cours du temps avec valuation constante.

L'algorithme *int* permet de lisser l'estimation en gardant une variabilité dans l'estimation. Finalement on observe bien la convergence prouvée par la loi forte des grands nombres pour l'algorithme *id* vers la consommation réelle.

La figure 5.3(a) représente l'erreur d'estimation sur le délai moyen pour chacun des algorithmes. Encore, l'algorithme *last* montre une grande variabilité dans l'estimation alors que l'algorithme *id* converge. On observe le même phénomène pour l'évolution du paramètre de mesure passive  $\epsilon$  sur la figure 5.3(b). De plus, on remarque que nous avons  $\gamma \ll \lambda$  et donc l'hypothèse d'indépendance du délai entre les sondes est valable.





(a) Évolution de l'erreur d'estimation du délai moyen pour chaque algorithme.

(b) Évolution du paramètre de mesures passives  $\epsilon$  pour chaque algorithme.

FIG. 5.3 – Évolution des paramètres de mesure actives et passives au cours du temps avec valuation constante.

### 5.6.2 Valuation variable

On suppose maintenant les fonctions de valuation suivantes :

$$\begin{aligned} V_1(\lambda_1, t) &= 15\lambda_1 - 3\lambda_1^2, \\ V_2(\lambda_2, t) &= 17\lambda_2 - 4\lambda_2^2, \\ V_3(\lambda_3, t) &= (16\lambda_3 - 5\lambda_3^2) \frac{t}{100}. \end{aligned}$$

Ainsi, la valuation pour les utilisateurs de classe 3 augmente progressivement dans le temps. Les figures 5.4(a), 5.4(b) et 5.4(c) représentent l'évolution des débits, de leurs estimations et du taux d'envoi des sondes  $\gamma$  pour chaque algorithme. Nous remarquons également que la variabilité des estimations est la plus grande pour l'algorithme *last* et la plus faible pour l'algorithme *id*.

Une petite variabilité en utilisant de façon identique les estimations du passé dans l'algorithme *id* induit un biais dans l'estimation. Ainsi, l'algorithme *int* (donné sur la figure 5.4(b)) avec les poids  $r_\lambda = r_d = 0.38$  est un bon compromis entre la variabilité et le biais d'estimation. La fonction de valuation  $V_3$  pour les utilisateurs de la troisième classe de trafic induit un débit variable qui croît constamment.

La figure 5.5(a) montre l'évolution de l'erreur d'estimation du délai moyen pour chaque algorithme. A cause du biais, l'algorithme *id* induit une mauvaise estimation, qui ne converge pas vers 0 comme avec demande constante, et même on observe que l'erreur augmente avec le temps. Ceci s'explique par le fait que sur la figure 5.4(c) on remarque que l'écart entre l'estimateur  $\hat{\lambda}_3$  et la consommation réelle augmente avec le temps. L'algorithme *id* n'arrive pas à rattraper le vrai débit à cause du poids du passé trop important. Notons que sur l'ensemble des itérations sont :

- pour l'algorithme *last*, la moyenne de l'erreur d'estimation du délai est 0.0251,

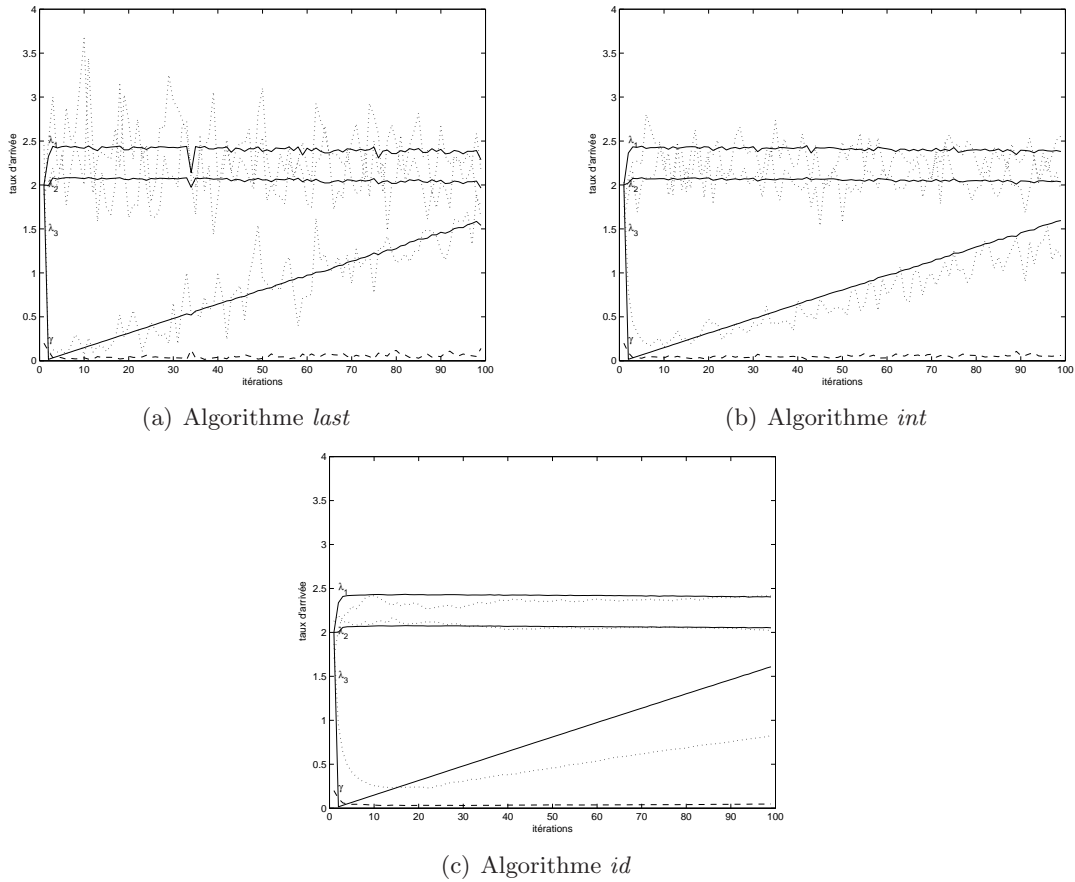


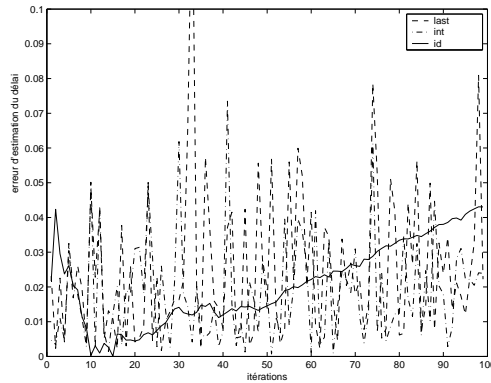
FIG. 5.4 – Évolution des débits  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  et  $\gamma$  au cours du temps avec valuation variable.

- pour l'algorithme *int* avec  $r_\lambda = r_d = 0.38$ , la moyenne de l'erreur d'estimation du délai est 0.0183,
- pour l'algorithme *id*, la moyenne de l'erreur d'estimation du délai est 0.0208.

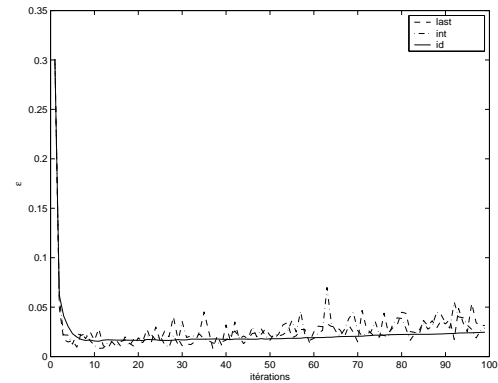
L'algorithme *int* est donc un bon compromis entre les algorithmes *last* et *id*. Nous avons proposé les paramètres  $r_\lambda = r_d = 0.38$  après de multiples manipulations. Enfin, la figure 5.5(b) présente l'évolution du paramètre de mesure passives  $\epsilon$  dans le temps. Celui-ci augmente car le débit total croît progressivement dans le temps à cause de  $\lambda_3$ .

## 5.7 Conclusion

Dans ce chapitre, nous avons supposé, dans un contexte de tarification à l'usage, que le système doit mesurer la consommation des utilisateurs de son réseau. En effet, le débit optimal  $\lambda^*$  induit par les prix et la relation de demande, n'est pas forcément le débit réel. Nous proposons alors d'utiliser des techniques provenant de la métrologie



(a) Évolution de l'erreur d'estimation du délai moyen pour chaque algorithme.



(b) Évolution du paramètre de mesures passives  $\epsilon$  pour chaque algorithme.

FIG. 5.5 – Évolution des paramètres de mesure actives et passives au cours du temps avec valuation variable.

des réseaux qui permettent d'obtenir des estimations des métriques comme le débit ou le délai de bout en bout. Le fournisseur détermine le prix pour chaque type de trafic afin d'optimiser le surplus collectif du système. Seulement, ces prix optimaux dépendent des débits mais également du délai moyen, lesquels sont alors estimés par le fournisseur. Nous construisons alors deux estimateurs sans biais afin de mesurer la consommation des utilisateurs et le délai moyen du système. Tout d'abord, nous utilisons des mesures actives avec envoi de sonde dans le système afin d'estimer le délai moyen. Puis, nous considérons des mesures passives d'observation du trafic pour estimer les débits. Nous déterminons alors explicitement les prix optimaux qui optimisent le surplus collectif du système pour ce modèle. Nous étudions ensuite l'optimisation des paramètres de mesures et proposons plusieurs algorithmes afin d'obtenir de meilleurs estimateurs des métriques suivant la fonction de valuation. Ces algorithmes sont comparés numériquement avec des comportements différents d'utilisateurs où la demande est constante ou variable dans le temps.



## Chapitre 6

# Tarification dans un réseau sans fil CDMA

Les études précédentes ont traité de réseaux filaires, mais le problème de congestion est d'autant plus important dans les réseaux sans fil où la capacité de transmission des paquets est limitée par la bande de fréquences. Le chapitre porte sur le contrôle de congestion via une tarification statique dans une cellule sans fil utilisant une technique de transmission CDMA [113] ( pour *Code Division Multiple Acces*).

L'apparition dans la téléphonie de troisième génération de nouvelles applications très demandeuses en ressource comme la télévision, la vidéo conférence, etc, induit d'important problème de congestion. Dans ce contexte, nous proposons un mécanisme de tarification à l'usage afin de contrôler la demande. On s'intéresse particulièrement aux transmissions sur la voie montante (transfert de fichiers), mobile vers station de base, car la capacité de transmission proposé par les normes standards sur la voie montante est moins importante que sur la voie descendante [55]. De plus, les codes de Walsh utilisés en CDMA proviennent des matrices d'Hadamard et l'orthogonalité entre deux codes différents n'est pas conservée si on décale un code par rapport à l'autre [21]. Donc la désynchronisation des signaux des mobiles à la station de base est due à la répartition géographique aléatoire des mobiles dans la cellule [56]. Ainsi, la capacité de transmission sur la voie descendante est plus importante que sur la voie montante.

La téléphonie de troisième génération (3G) a été standardisée par l'ETSI (European Telecommunication Standard Institute) sous le nom de réseaux UMTS (pour *Universal Mobile Telecommunication System*). L'UMTS utilise, sur l'interface radio, la technique CDMA comme mode d'accès à utilisateurs multiples. Cette technique permet de desservir plusieurs utilisateurs sur la même bande de fréquence et pendant le même slot de temps grâce à des codes orthogonaux. Ainsi, le CDMA permet au réseau UMTS d'obtenir des débits importants et ce type de réseau sans fil propose de nombreuses classes d'applications temps-réel comme la visiophonie, l'accès à l'Internet, la télévision, etc. La qualité de service d'un mobile dépend de son niveau d'interférence et ainsi, des puissances d'émission de l'ensemble des autres mobiles de la cellule et des cellules voisines. Nous supposons que le contrôle de puissance est parfait (*perfect power control*) dans

le sens où l'ensemble des utilisateurs d'une même classe de service possède la même puissance de réception à la station de base. Dans [10], les auteurs proposent un algorithme à pas fixe afin d'obtenir le contrôle de puissance parfait. Cette hypothèse est importante dans le cas des transmissions sur la voie montante car elle permet d'éliminer les effets «proche-lointain» lorsqu'un utilisateur avec une puissance importante proche de la station de base masque d'autres avec une plus faible puissance qui transmettent de plus loin [48, 103].

Il existe de nombreux articles traitant du problème d'allocation de puissance et de gestion de la qualité de service dans ce type de réseau (voir [69, 99]). Ces modèles tentent d'obtenir la meilleure qualité de service (niveau d'interférences) pour les utilisateurs présents dans le système mais ne cherchent pas à contrôler la demande. Pour cela, l'utilisation de la tarification à l'usage où les prix dépendent de la puissance d'émission a été proposée dans [77, 104] par exemple.

Dans notre modèle, le système propose une tarification statique à l'usage des services dans une cellule CDMA contrairement aux modèles de la littérature qui considèrent généralement une tarification dynamique. En effet, dans la plupart des modèles comme dans [42], le prix dépend de la puissance d'émission qui varie suivant la distance du mobile à la station de base afin d'avoir un niveau de qualité satisfaisant. Nous considérons une tarification statique qui est mieux appréciée des utilisateurs car les prix sont fixés [41]. Nous déterminons alors les prix et les puissances (de réception car nous avons un contrôle de puissance parfait) qui optimisent le revenu de la station de base. Le modèle de réseau s'inspire de [61], où une allocation optimale de ressource est obtenue pour plusieurs classes de trafic, mais pour un nombre fixé d'utilisateurs dans chaque classe. Dans cet article, la puissance est contrôlée afin d'obtenir des seuils du rapport d'interférences signal à bruit (SINR pour *Signal to Interference plus Noise Ratio*) pour lesquels un niveau satisfaisant de qualité de service est atteint. Nous introduisons une contrainte de capacité afin de limiter le nombre d'utilisateurs dans une cellule en fonction de la puissance de réception. Cette contrainte est définie dans [113]. Nous étudions un modèle de tarification à l'usage avec une ou deux différentes classes d'applications ou de trafic. Nous supposons tout d'abord que la demande excède toujours la capacité, puis nous étudions le cas où la demande est aléatoire, représentant par exemple le comportement hétérogène des utilisateurs sur un jour entier. Le nombre d'utilisateurs voulant accéder aux services varie entre la nuit et les heures de pointe.

Nous introduisons tout d'abord notre modèle de réseau sans fil avec une cellule utilisant la technologie CDMA. Nous considérons uniquement les interférences intra-cellulaires, les interférences inter-cellulaires étant prises en compte dans le bruit. Nous présentons ensuite notre modèle de tarification à l'usage afin de contrôler la demande dans ce type de réseau.

## 6.1 Modèle de transmission CDMA

Le modèle de transmission de paquets d'un utilisateur de type  $i$  vers la station de base (voie montante) dans une cellule CDMA est représenté figure 6.1.

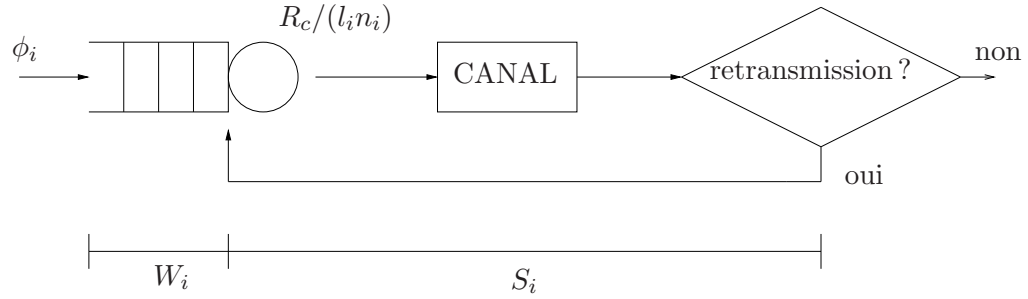


FIG. 6.1 – Modèle de transmission du trafic de type  $i$  sur la voie montante dans une cellule CDMA.  $W_i$  est le délai moyen dans la file d’attente et  $S_i$  est le temps moyen de transfert incluant les retransmissions.

Chaque paquet transmis comporte des successions de groupe de bits de données appelées symboles. Nous supposons que le débit des symboles  $R_c$ , qui définit le nombre de paquets codés par seconde, est le même pour toutes les classes d’utilisateurs. Les paquets à coder sont envoyés à travers un buffer modélisant la conversion du signal (des symboles) en code CDMA avec un débit de conversion  $R_c/(l_i n_i)$  où  $l_i$  le nombre de symboles et  $n_i$  est le facteur d’étalement pour un paquet de classe  $i$  en nombre de paquets par symboles. En supposant que les paquets à coder entrent dans la file avec un taux  $\phi_i$ , le facteur d’étalement  $n_i$  doit être plus petit que  $R_c/(l_i \phi_i)$  afin d’éviter la saturation du buffer de codage. Nous considérons qu’un nouveau paquet est généré (entre dans la file du codeur) dès que le précédent est envoyé avec succès sur le canal. Ainsi, tous les utilisateurs présents dans la cellule sont supposés être actifs en permanence, c’est-à-dire qu’ils envoient sans arrêt leur paquet sur le canal. Ce modèle de génération des paquets peut représenter la transmission des longs fichiers [61] par exemple. On suppose un contrôle parfait de la puissance. La station de base envoie aux mobiles de classe  $i$  des messages de contrôle afin que la puissance reçue par la station de base d’un mobile de la classe  $i$  soit égale à  $P_i$ . Ainsi  $P_i$  est la puissance de réception à la station de base associée à tous les utilisateurs de classe  $i$ .

### 6.1.1 Mesures de performance

Dans les transmissions sans fil, une quantité importante qui détermine la qualité de service est le rapport puissance sur les interférences plus bruit ( $SINR$ ). En effet, les mesures de performances telles le délai et la probabilité d’erreur d’un bit s’exprime explicitement à partir de cette quantité. Pour les paquets de classe  $i$ , nous avons

$$SINR_i = \frac{P_i n_i}{\nu \left( (N_i - 1) P_i + \sum_{j \neq i}^K N_j P_j \right) + \chi^2}, \quad (6.1)$$

où  $\nu$  est une constante qui dépend de la forme du chip,  $N_j$  est le nombre d’utilisateurs de classe  $j$  et  $\chi^2$  est la puissance du bruit qui inclue les interférences inter-cellulaires.

Pour chaque type ou classe de trafic, nous supposons comme dans [61] que la probabilité d'erreur d'un bit (BEP pour *Bit Error Probability*) est une fonction exponentielle décroissante du *SINR*. Spécifiquement, nous supposons que pour un utilisateur de type  $i$ , le BEP est

$$p_{b_i} = \mathcal{F}(\text{SINR}_i), \quad (6.2)$$

avec  $\mathcal{F}(x) = \exp(-\omega x)$ . Dans [61], les auteurs utilisent une fonction  $\mathcal{F}(x) = \kappa \exp(-\omega x)$  avec  $\kappa \neq 1$ . Seulement le délai de transmission est alors borné lorsque la puissance est réduite à zéro. Afin de prévenir ce problème de modélisation, nous considérons  $\kappa = 1$  comme dans [42] et ainsi le délai est infini et la probabilité de retransmission égale à 1 lorsque la puissance est nulle. La probabilité de retransmission s'exprime par

$$p_{r_i} = 1 - [1 - \mathcal{F}(\text{SINR}_i)]^{l_i r_i} \quad (6.3)$$

où  $r_i$  est le taux de redondance utilisé par le mécanisme de FEC. Le délai moyen  $\mathbb{E}D_i$  d'un paquet de classe  $i$  s'exprime comme la somme du délai dans la file d'attente  $W_i$  et du délai de retransmission  $S_i$  (voir figure 6.1). Il a été montré dans [61] que

$$\mathbb{E}D_i = \frac{l_i n_i}{R_c(1 - p_{r_i})} = \frac{l_i n_i}{R_c(1 - \exp(-\omega \text{SINR}_i))^{l_i r_i}}. \quad (6.4)$$

### 6.1.2 Contrainte de capacité

Le délai moyen augmente avec le nombre d'utilisateurs dans la cellule. Plus le nombre d'utilisateurs augmente et plus les interférences vont augmenter également. Afin de limiter le nombre d'utilisateurs, nous utilisons la contrainte de capacité décrite dans [113] par

$$\frac{\nu \left( (N_i - 1)P_i + \sum_{j \neq i}^K N_j P_j \right) + \chi^2}{\chi^2} \leq \frac{1}{\eta} \quad (6.5)$$

où  $\eta$  est typiquement compris entre 0.1 et 0.25. La puissance du bruit  $\chi^2$  doit être égale à au moins 25% de l'interférence totale, ce qui limite le nombre de mobiles. Cette inégalité donne une borne supérieure sur le nombre d'utilisateurs de chaque type présents dans la cellule en fonction des puissances.

## 6.2 Modélisation des utilisateurs

Dans la suite de ce chapitre, nous notons par  $i \in \{v, d\}$  la classe de trafic, avec  $v$  pour le trafic de type «voix» et  $d$  pour «données». Le trafic de type «voix» ne correspond pas à un trafic réellement multimédia, nous choisissons cette dénomination afin de le différencier avec le trafic de type «données» qui est supposé moins sensible au niveau de qualité de service. Lorsque l'on considère une unique classe de trafic, nous ôtons l'indice  $i$ .



L'utilité  $U_i$  des utilisateurs de type  $i$  s'exprime comme la différence entre la valuation du niveau de qualité de service (fonction du SINR, lui-même fonction du nombre d'utilisateurs) et du prix par paquet  $u_i$ . Nous posons

$$\forall i \in \{v, d\} \quad U_i(N_v, N_d) = V_i(\text{SINR}_i(N_v, N_d)) - u_i. \quad (6.6)$$

La fonction  $V_i$  décrit la valuation du niveau de qualité de service perçu par un utilisateur de type  $i$  par rapport à son SINR. Cette fonction vérifie les hypothèses suivantes.

**Hypothèse 6** *Pour tout  $i \in \{v, d\}$ , la fonction de valuation  $V_i$  est strictement croissante, dérivable et vérifie  $V_i(0) = 0$ .*

Lorsque le rapport puissance sur interférence est nul, alors la valuation du service est nulle. Et, plus la puissance est grande par rapport aux interférences, ce qui implique une augmentation du SINR, plus la valuation du service perçu augmente.

Prenons par exemple la fonction d'utilité du chapitre 3 qui dépend du délai moyen par :

$$U_i(N_v, N_d) = \frac{1}{(\text{ED}_i(N_v, N_d))^{\alpha_i}} - u_i, \quad (6.7)$$

où  $\alpha_i$  est la sensibilité du trafic de classe  $i$  face au délai moyen. D'après l'équation (6.4), nous obtenons la fonction de valuation

$$\forall i \in \{v, d\}, \quad V_i(x) = \left( \frac{R_c}{l_i r_i} \right)^{\alpha_i} (1 - e^{-\omega x})^{l_i r_i \alpha_i},$$

qui vérifie bien l'hypothèse 6. Ainsi les résultats obtenus dans la suite avec une fonction d'utilité générale, restent valides avec la fonction d'utilité exprimée par l'équation (6.7).

### 6.3 Jeu non-coopératif entre classes de service

Les utilisateurs de classe  $i$  demande à rejoindre le système tant que leur utilité  $U_i$  est positive. On dit alors que les utilisateurs possèdent un intérêt à rejoindre le système. Au contraire, si leur utilité est négative, les utilisateurs quittent le système. La fonction d'utilité dépend explicitement du nombre d'utilisateurs  $N_v$  et  $N_d$ . Ainsi, pour des prix et des puissances fixés, nous avons, comme dans le chapitre 3.4 un jeu non-coopératif entre les classes de service induit par la décision de chaque utilisateur de rejoindre ou quitter le système. La théorie des jeux est un outil largement utilisé pour les problèmes de contrôle de puissance dans les réseaux sans fil [5, 42]. Nous déterminons l'existence et l'unicité d'un équilibre de Nash pour ce jeu, lequel correspond à une situation de stabilisation du nombre d'utilisateurs de chaque type.

Pour chaque classe  $i \in \{v, d\}$ , le nombre d'utilisateurs  $N_i^*$  à l'équilibre est décrit de la façon suivante.

1. Le nombre d'utilisateurs est nul avec une utilité négative, ce qui représente le fait que aucun utilisateur possède un intérêt à rejoindre le système, donc

$$N_i^* = 0 \quad \text{et} \quad \forall N_j > 0 \quad U_i(1, N_j) < 0.$$

2. Le nombre d'utilisateurs annule la contrainte de capacité de la cellule avec une utilité positive, et donc, aucun autre utilisateur est autorisé à envoyer ses paquets. Nous avons donc

$$N_i^* \text{ annule la contrainte donnée par l'équation (6.5) et } U_i(N_v^*, N_d^*) > 0.$$

3. Enfin, le nombre d'utilisateurs est positif et inférieur à la contrainte de capacité avec une utilité nulle, ainsi aucun autre utilisateur possède un intérêt à envoyer ses paquets car cela rendrait son utilité négative. Nous avons donc

$$N_i^* > 0 \text{ vérifie la contrainte donnée par l'équation (6.5) et } U_i(N_v^*, N_d^*) = 0.$$

Nous nous intéressons maintenant à déterminer l'existence de l'équilibre en fonction des prix et des puissances puis nous optimisons le revenu de la station de base en fonction de ces variables. On étudie tout d'abord, pour simplifier l'analyse et mieux faire comprendre le fonctionnement, le cas d'un unique type de trafic.

## 6.4 Tarification optimale avec une classe de service

On considère un unique type de trafic et pour simplifier les notations, nous ôtons les indices  $v$  et  $d$ . Nous supposons dans un premier temps que la demande excède toujours la capacité et est donc infinie. Nous étudions ensuite le cas plus réaliste de demande aléatoire, où le nombre d'utilisateurs voulant rejoindre le système est représenté par une variable aléatoire discrète. Elle modélise par exemple la variation du nombre d'utilisateur demandant l'accès à la station de base à différentes périodes de la journée. Il est alors évident que la demande est différente entre certains moments de la journée et elle peut donc être inférieure à la capacité à certains instants.

### 6.4.1 Demande infinie

Nous considérons tout d'abord le cas d'une fonction  $V_i$  générique qui vérifie l'hypothèse 6. Ensuite, nous appliquons nos résultats au cas particulier d'une fonction d'utilité qui dépend du délai moyen comme exprimée par la relation (6.7).

#### 6.4.1.1 Utilité générale

Nous supposons que la fonction d'utilité  $U$  s'exprime en fonction du nombre d'utilisateur  $N$  et du prix par paquet  $u$  par

$$U(N) = V(\text{SINR}(N)) - u,$$

où la fonction de valuation  $V$  est strictement croissante, dérivable et vérifie  $V(0) = 0$  (hypothèse 6). Le théorème suivant donne l'expression du nombre d'utilisateurs à l'équilibre en fonction du prix  $u$  et de la puissance  $P$ .

**Théorème 14** Soient  $u$  le prix par paquet et  $P$  la puissance de réception à la station de base pour tous les utilisateurs. Le nombre d'utilisateurs  $N^*$  à l'équilibre du système est :

$$N^* = \begin{cases} 0, & \text{si } u > u_{max} = V(\frac{nP}{\chi^2}), \\ 1 - \frac{\chi^2}{\nu P} - \frac{n}{\nu V^{-1}(u)}, & \text{si } V(\frac{nP\eta}{\chi^2}) < u < V(\frac{nP}{\chi^2}), \\ 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P}, & \text{sinon.} \end{cases}$$

Le premier cas correspond à la situation où l'accès au service est trop cher pour les utilisateurs et donc la quantité  $V(\frac{nP}{\chi^2})$  représente bien un prix maximum. Quant au dernier, il représente le cas où la capacité maximum en fonction de la puissance est atteinte et donc la contrainte de capacité donnée par l'équation (6.5) est active.

**Preuve** Nous considérons que le prix par paquet  $u$  et la puissance de réception à la station de base  $P$  sont fixés. L'objectif est de déterminer le nombre d'utilisateurs  $N^*$  tel que aucun autre utilisateur ne possède un intérêt à rejoindre ou quitter le système (équilibre de Nash). Nous cherchons  $N^*$  tel que

1.  $U(1) \leq 0$  et  $N^* = 0$ ,
2. ou  $U(N^*) > 0$  et  $N^* = \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P} + 1$  (cette dernière égalité correspond au cas où la contrainte de capacité (6.5) est atteinte),
3. ou  $U(N^*) = 0$  et  $0 < N^* < \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P} + 1$ .

Nous avons

$$\begin{aligned} U(N^*) = 0 & \Leftrightarrow V(SINR(N^*)) = u, \\ & \Leftrightarrow SINR(N^*) = \frac{Pn}{\nu(N^* - 1)P + \chi^2} = V^{-1}(u). \end{aligned}$$

Comme le  $SINR$  est une fonction strictement décroissante en  $N$ , cette équation possède une solution, alors unique, si

1.  $SINR(1)$  correspondant au SINR maximum est supérieur à  $V^{-1}(u)$  et
2. le  $SINR$  pour la contrainte de capacité est inférieur à  $V^{-1}(u)$ .

La première condition,  $SINR(1) > V^{-1}(u)$ , définit le prix maximum

$$u_{max} = V(SINR(1)) = V(nP/\chi^2),$$

au-dessus duquel aucun utilisateur rejoint le système car leur utilité est négative et donc  $N^* = 0$ . La seconde, liée à la contrainte de capacité, définit un prix  $V(SINR(\frac{1-\eta}{\eta} \frac{\chi^2}{\nu P} + 1)) = V(\frac{nP\eta}{\chi^2})$  au-dessus duquel le nombre d'utilisateurs est inférieur à la capacité. Si  $V(\frac{nP\eta}{\chi^2}) < u < u_{max}$ , le nombre d'utilisateurs  $N^*$  à l'équilibre vérifie

$$V(SINR(N^*)) = u \quad \Leftrightarrow \quad N^* = 1 - \frac{\chi^2}{\nu P} - \frac{n}{\nu V^{-1}(u)}.$$

■

Le revenu de la station de base à l'équilibre du système est égal à

$$R(u, P) = uN^*(u, P)\lambda(u, P),$$

avec  $\lambda(u, P)$  le débit moyen d'envoi des paquets à la station de base et  $N^*(u, P)$  le nombre d'utilisateurs à l'équilibre donné par le théorème précédent. Le débit moyen représente le nombre moyen de bits transmis avec succès et s'exprime en fonction du délai moyen par

$$\lambda(u, P) = \frac{Lr}{\mathbb{E}D(u, P)} = \frac{rR_c}{n} \left( 1 - \exp\left(-\omega \frac{Pn}{\nu(N^*(u, P) - 1)P + \chi^2}\right) \right)^{lr}. \quad (6.8)$$

Le théorème suivant donne l'expression du prix et de la puissance qui optimisent le revenu de la station de base à l'équilibre du système.

**Théorème 15** Soit  $\mathcal{X}$  l'ensemble des solutions de l'équation suivante sur  $X > 0$  :

$$\begin{aligned} V(X) (\nu X^2 lr \omega e^{-\omega X} + Xn(1 - \eta)lr\omega - n(1 - e^{-\omega X})) \\ + f'(X) (\nu X^2(1 - e^{-\omega X}) + Xn(1 - e^{-\omega X})) = 0. \end{aligned} \quad (6.9)$$

Soit

$$X^* = \arg \max_{X \in \mathcal{X}} \left( \frac{rR_c}{n} (1 - e^{-\omega X})^{lr} V(X) + \frac{1 - \eta}{\nu} rR_c \frac{(1 - e^{-\omega X})^{lr} V(X)}{X} \right).$$

Alors, le prix par paquet  $u^*$  et la puissance de réception  $P^*$  qui maximisent le revenu de la station de base sont

$$u^* = V(X^*) \text{ et } P^* = \frac{X^* \chi^2}{n\eta}.$$

**Preuve** Le revenu de la station de base s'exprime en fonction du prix par paquet  $u$  et de la puissance de réception  $P$  par

$$R(u, P) = uN^*(u, P)\lambda(u, P).$$

Le théorème 14 donne trois expressions de  $N^*$  en fonction du couple  $(u, P)$  (correspondant aux trois domaines A, B et C représentés sur la figure 6.2).

- Sur le domaine B où  $V(\frac{nP\eta}{\chi^2}) < u < V(\frac{nP}{\chi^2})$  nous avons d'après le théorème 14,  $N^*(u, P) = 1 - \frac{\chi^2}{\nu P} - \frac{n}{\nu V^{-1}(u)}$ . Ainsi, le revenu de la station de base s'exprime explicitement en fonction du prix par paquet et de la puissance par

$$R(u, P) = u \left( 1 - \frac{\chi^2}{\nu P} - \frac{n}{\nu V^{-1}(u)} \right) \frac{rR_c}{n} \left( 1 - e^{-\omega V^{-1}(u)} \right)^{lr}. \quad (6.10)$$

Cette fonction est continue et dérivable en toutes ses variables et on peut vérifier que l'on a

$$\forall u \geq 0, P \geq 0, \quad \frac{\partial R}{\partial P}(u, P) = \frac{u\chi^2}{\nu P^2} > 0.$$

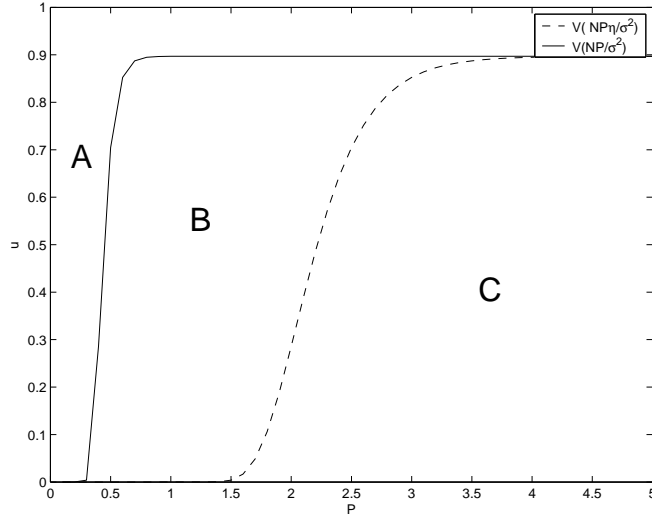


FIG. 6.2 – Domaines d'équilibre ("personne" zone A, "capacité atteinte" zone C ou "utilité nulle" zone B) du nombre d'utilisateurs en fonction du prix par paquet  $u$  et de la puissance de réception  $P$  pour un seul type de trafic.

Donc, sur le domaine B, le revenu est maximum si  $P = \frac{\chi^2}{n\eta} V^{-1}(u)$ , i.e.,  $u = V(\frac{nP\eta}{\chi^2})$  (courbe en pointillé sur la figure 6.2). Ainsi les prix et les puissances optimaux sur le domaine B sont tels que la capacité de la cellule est atteinte.

- Sur le domaine C où  $V(\frac{nP\eta}{\chi^2}) > u$ , le théorème 14 donne  $N^*(u, P) = 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P}$  et donc le revenu de la station de base est

$$R(u, P) = u \left( 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P} \right) \frac{rR_c}{n} \left( 1 - e^{-\omega \frac{Pn}{\chi^2} \eta} \right)^{lr}. \quad (6.11)$$

Cette fonction est également continue et dérivable en toutes ses variables et nous avons

$$\frac{\partial R}{\partial u}(u, P) = \left( 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P} \right) \frac{rR_c}{n} \left( 1 - e^{-\omega \frac{Pn}{\chi^2} \eta} \right)^{lr} > 0.$$

De même, le revenu est optimisé sur le domaine C en un point  $(u, P)$  tel que

$$u = V\left(\frac{nP\eta}{\chi^2}\right).$$

- Sur le domaine A, nous avons  $N^* = 0$  et donc le revenu est nul.

D'après les résultats précédents, le revenu de la station de base est maximisé sur la courbe  $u = V(\frac{nP\eta}{\chi^2})$ , où le revenu s'exprime en fonction de la puissance de réception  $P$  par

$$\begin{aligned} R(P) &= R\left(V\left(\frac{nP\eta}{\chi^2}\right), P\right), \\ &= V\left(\frac{nP\eta}{\chi^2}\right) \left( 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P} \right) \frac{rR_c}{n} \left( 1 - e^{-\omega \frac{Pn}{\chi^2} \eta} \right)^{lr}. \end{aligned}$$

Cette fonction est continue et dérivable sur  $[0, +\infty[$ . Utilisons le changement de variables suivant :  $X = \frac{nP}{\chi^2}\eta$ . Le revenu s'exprime en fonction de la variable  $X$  par

$$R_1(X) = R(P) = \frac{rR_c}{n}(1 - e^{-\omega X})^{lr}V(X) + \frac{1 - \eta}{\nu}rR_c \frac{(1 - e^{-\omega X})^{lr}V(X)}{X}.$$

L'équation  $\frac{\partial R}{\partial P}(P) = 0$  est équivalente à  $\frac{\partial R_1}{\partial X}(X) = 0$  car  $\frac{\partial X}{\partial P} = \frac{n}{\chi^2}\eta \neq 0$ . Or,

$$\begin{aligned} \frac{\partial R_1}{\partial X}(X) &= \frac{rR_c}{n} \left( lr\omega e^{-\omega X} (1 - e^{-\omega X})^{lr-1} V(X) + (1 - e^{-\omega X})^{lr} f'(X) \right) \\ &\quad + \frac{1 - \eta}{\nu} rR_c \frac{lr\omega X (1 - e^{-\omega X})^{lr-1} V(X) + X (1 - e^{-\omega X})^{lr} f'(X) - (1 - e^{-\omega X})^{lr} V(X)}{X^2}. \end{aligned}$$

Après simplifications nous obtenons  $\frac{\partial R_1}{\partial X}(X) = 0$  si et seulement si

$$\begin{aligned} V(X) (\nu X^2 lr\omega e^{-\omega X} + Xn(1 - \eta)lr\omega - n(1 - e^{-\omega X})) \\ + f'(X) (\nu X^2 (1 - e^{-\omega X}) + Xn(1 - e^{-\omega X})) = 0. \end{aligned}$$

Soit  $\mathcal{X}$  l'ensemble des solutions de cette équation et  $X^*$  la solution, si il en existe, qui optimise la fonction  $R_1$ , c'est-à-dire

$$X^* = \arg \max_{X \in \mathcal{X}} \left( \frac{rR_c}{n} (1 - e^{-\omega X})^{lr} V(X) + \frac{1 - \eta}{\nu} rR_c \frac{(1 - e^{-\omega X})^{lr} V(X)}{X} \right).$$

La puissance optimale est donc

$$P^* = \frac{X^* \chi^2}{n\eta},$$

et comme nous avons la relation  $u = V(\frac{NP\eta}{\chi^2})$ , nous obtenons le prix optimal

$$u^* = V\left(\frac{nP^*\eta}{\chi^2}\right) = V(X^*).$$

■

Dans la section suivante, nous proposons un corollaire de ce théorème pour un cas particulier de fonction d'utilité. Nous considérons la fonction qui dépend du délai moyen des paquets donnée par l'expression (6.7). Cette fonction a également été proposée dans le chapitre 3 et est utilisée dans de nombreuses analyses économiques (famille de fonctions de Cobb-Douglas).

#### 6.4.1.2 Utilité en fonction du délai moyen

Nous supposons maintenant que la valuation des utilisateurs dépend du délai moyen de transmission de leurs paquets. La fonction d'utilité s'exprime par

$$U(N) = \frac{1}{ED^\alpha} - u,$$

où  $\alpha$  est le paramètre de sensibilité des utilisateurs au délai moyen. Ainsi, la fonction de valuation  $V$  (pour le *SINR*) correspondante est

$$V(x) = \left(\frac{R_c}{ln}\right)^\alpha (1 - e^{-\omega x})^{lr\alpha}.$$

Cette fonction vérifie l'hypothèse 6 sur la fonction de valuation et nous pouvons donc appliquer directement les théorèmes 14 et 15. Nous déterminons ainsi le nombre d'utilisateurs à l'équilibre et optimisons le revenu de la station de base en fonction du prix et de la puissance. Ces résultats sont regroupés dans le corollaire suivant.

**Corollaire 1** *Nous supposons que la fonction d'utilité dépend du délai moyen et s'exprime par  $U(N) = \frac{1}{ED^\alpha(N)} - u$ . Alors*

– le nombre d'utilisateurs  $N^*$  à l'équilibre en fonction de  $(P, u)$  est

$$N^* = \begin{cases} 0, & \text{si } u > u_{max} = \left(\frac{R_c}{ln}\right)^\alpha (1 - e^{-\omega \frac{nP}{x^2}})^{lr\alpha}, \\ 1 - \frac{\chi^2}{\nu P} - \frac{\omega n}{\nu \ln(1 - \left(\frac{ln}{R_c} u^{1/\alpha}\right)^{1/lr})}, & \text{si } \left(\frac{R_c}{ln}\right)^\alpha (1 - e^{-\omega \frac{nP\eta}{x^2}})^{lr\alpha} < u \\ 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P}, & \text{et } u < \left(\frac{R_c}{ln}\right)^\alpha (1 - e^{-\omega \frac{nP}{x^2}})^{lr\alpha}, \\ & \text{sinon.} \end{cases}$$

– Soit  $X^*$  l'**unique** solution positive de l'équation suivante

$$lr(1 + \alpha)\omega X^2 + \frac{1 - \eta}{\nu} lr(1 + \alpha)\omega n X + (1 - \eta)\frac{n}{\nu} = \frac{1 - \eta}{\nu} ne^{\omega X}.$$

Le prix par paquet optimal est

$$u^* = \left(\frac{R_c}{LN}\right)^\alpha (1 - e^{-\omega X^*})^{\alpha Lr},$$

est la puissance de réception optimale est

$$P^* = \frac{X^* \chi^2}{n\eta}.$$

Afin de prouver ce corollaire, nous avons besoin du lemme suivant.

**Lemme 8** *Soient  $a$  et  $b$  deux réels tels que  $a > 0$  et  $b > 1$ . L'équation suivante*

$$aX^2 + bX + 1 = e^X,$$

*possède une unique solution positive.*

**Preuve** Soit la fonction  $F(X) = aX^2 + bX + 1 - e^X$ . Nous avons  $F(0) = 0$ ,  $\lim_{X \rightarrow \infty} F(X) = -\infty$ ,  $F'(X) = 2aX + b$  avec  $F'(0) = b - 1 > 0$ , et  $F''(X) = 2a - e^X$ .

- Si  $2a < 1$ ,  $F'$  est strictement décroissante sur  $[0, \infty)$  (car  $F'' < 0$ ). Comme  $F'(0) > 0$ , il existe alors une unique solution positive à l'équation.
- Si  $2a \geq 1$ ,  $F'$  est croissante car  $F'(0) > 0$  puis décroissante vers  $-\infty$  lorsque  $X \rightarrow \infty$ . Donc  $F$  est tout d'abord croissante positive car  $F(0) = 0$ , puis décroissante vers  $-\infty$ . Ainsi, il existe une unique solution positive à l'équation  $F(X) = 0$ . ■

Nous donnons maintenant la preuve du corollaire.

**Preuve du corollaire** En considérant la fonction d'utilité (6.7), l'équation (6.9) devient

$$lr(1 + \alpha)\omega X^2 + \frac{1 - \eta}{\nu} lr(1 + \alpha)\omega nX + (1 - \eta)\frac{n}{\nu} = \frac{1 - \eta}{\nu} ne^{\omega X}, \quad (6.12)$$

soit

$$aY^2 + bY + 1 = e^Y,$$

avec  $Y = \omega X$ ,  $a = lr(1 + \alpha)\frac{\nu}{(1 - \eta)n\omega}$  et  $b = lr(1 + \alpha)$ . Nous avons  $a > 0$  et  $b > 1$  car  $b = lr(1 + \alpha) > lr$  et  $lr$  est le nombre d'octets par paquet lequel est typiquement plus grand que 1. D'après le lemme 8, l'équation (6.12) possède une unique solution  $X^*$  sur  $X > 0$ . Ainsi, le théorème 15 donne le prix par paquet optimal

$$u^* = \left(\frac{R_c}{ln}\right)^\alpha (1 - e^{-\omega X^*})^{\alpha lr},$$

et la puissance de réception optimale  $P^* = \frac{X^* \chi^2}{n\eta}$ . ■

*Exemple* Nous illustrons maintenant notre résultat par une application numérique. Soient les variables suivantes :  $l = 768$  symboles,  $r = 1/2$ ,  $\omega = 2$ ,  $\nu = 2/3$  correspondant au chip rectangulaire,  $R_c = 5$  Mchips/s,  $\chi^2 = 0.9$  dB,  $\alpha = 1.6$  et  $n = 10$ . Nous considérons également la borne supérieure pour la contrainte de capacité,  $\eta = 0.2$ . Le corollaire 1 donne le prix par paquet optimal  $u^* = 0.4804$  et la puissance de réception optimale  $P^* = 2.1352$ . Le revenu maximum est alors  $R^* = R(u^*, P^*) = 411.699$ . Le revenu est représenté sur la figure 6.3 en fonction du prix par paquet  $u$  et la puissance  $P$ . On observe que nos résultats théoriques sont bien en accord avec le résultat numérique.

Considérons maintenant le cas plus réaliste où la demande (nombre d'utilisateurs voulant joindre le système) est représentée par une variable aléatoire discrète positive. Ainsi, le système détermine le prix par paquet et la puissance de réception qui optimisent son revenu moyen à long terme.

## 6.4.2 Demande aléatoire

Typiquement, le nombre d'utilisateurs qui demande à transmettre des paquets n'est pas constant, et peut être inférieure à la capacité à certains moments de la journée par exemple. Nous supposons que le nombre d'utilisateurs est représenté par une variable aléatoire positive discrète  $\Delta$ . L'objectif de la station de base est de déterminer le prix par



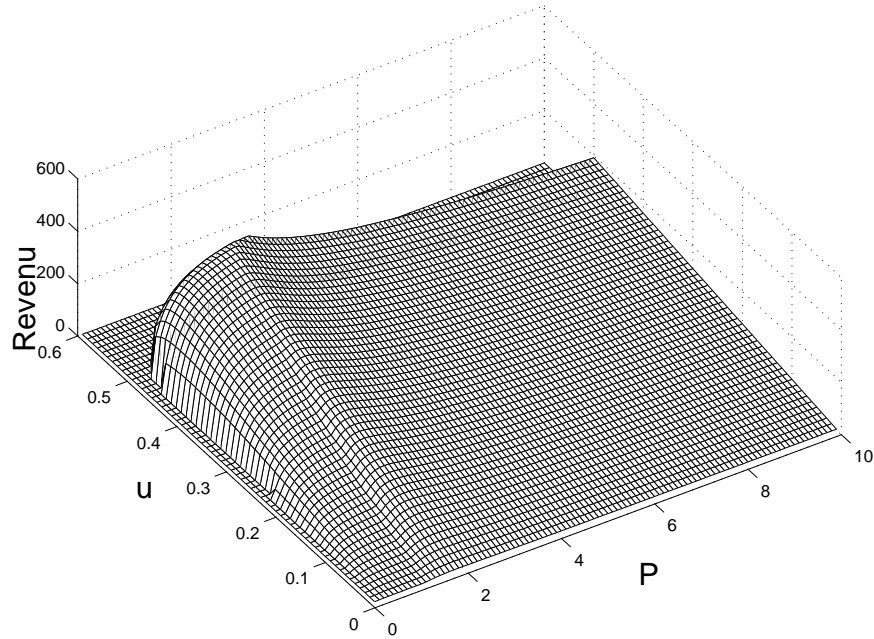


FIG. 6.3 – Revenu de la station de base en fonction du prix par paquet et de la puissance de réception lorsque la demande excède toujours la capacité et l'utilité dépend du délai moyen.

paquet  $u$  et la puissance  $P$  qui maximisent son revenu moyen. Tout d'abord, comme pour le cas avec demande supérieure à la capacité, nous déterminons le nombre  $\kappa^*$  d'utilisateurs à l'équilibre du système en fonction du prix  $u$  et de la puissance  $P$ .

**Théorème 16** Soient  $u$  le prix par paquet,  $P$  la puissance de réception à la station de base et  $\delta$  le nombre d'utilisateurs demandant le service à un certain instant. Le nombre d'utilisateurs  $\kappa^*$  à l'équilibre est défini par

$$\kappa^* = \min(\delta, N^*),$$

avec  $N^*$  donné par le théorème 14.

**Preuve** On considère le prix par paquet  $u$  et la puissance de réception  $P$  fixés. Suivant demande à un instant donné  $\delta$ , nous avons différents équilibres.

- Si  $\delta = 0$ , la demande est nulle et donc le nombre d'utilisateurs à l'équilibre est  $\kappa^* = 0$ .
- Si  $\delta > N^*$  où  $N^*$  est le nombre d'utilisateurs à l'équilibre donné par le théorème 14, alors l'équilibre est  $\kappa^* = N^*$ . En effet, malgré que la demande soit supérieure à  $N^*$ , il ne peut y avoir plus de  $N^*$  utilisateurs dans le système car soit l'utilité est négative ou la capacité est atteinte. Donc l'équilibre est  $\kappa^* = N^*$ .

- Enfin, si  $\delta \leq N^*$ , tous les utilisateurs jusqu'à la demande totale  $\delta$  sont servis et donc l'équilibre est  $\kappa^* = \delta$ .

Ainsi, nous résumons l'ensemble de ces situations par  $\kappa^*$  égale à

$$\kappa^* = \min(\delta, N^*),$$

avec  $N^*$  donné par le théorème 14. ■

Le revenu moyen de la station de base est défini en fonction du prix  $u$  et de la puissance  $P$  par

$$\bar{R}(u, P) = \sum_{\delta=0}^{\infty} u\kappa^*(u, P, \delta)\lambda(u, P, \delta)\mathbb{P}(\Delta = \delta), \quad (6.13)$$

où  $\kappa^*(u, P, \delta)$  est le nombre d'utilisateurs à l'équilibre donné par le théorème précédent, et  $\lambda(u, P, \delta)$  le débit moyen d'envoi des paquets à la station de base.

*Exemple* Supposons que la demande  $\Delta$  suit une loi de Poisson de taux 3 et considérons les paramètres de l'application numérique précédente. Le revenu moyen de la station de base est représenté sur la figure 6.4 et nous obtenons les valeurs du prix par paquet optimal  $u^* = 0.4905$  et de la puissance de réception optimale  $P^* = 2.2694$  pour une revenu moyen maximum  $R^* = 299.2589$ .

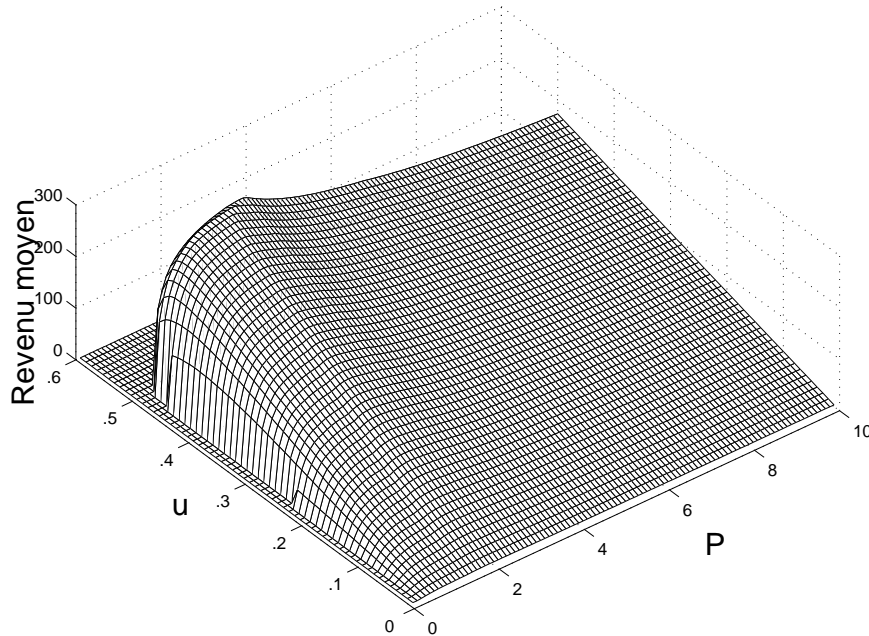


FIG. 6.4 – Revenu moyen de la station de base en fonction du prix par paquet et de la puissance de réception lorsque la demande est aléatoire et l'utilité dépend du délai moyen.

L'une des avancées technologiques importante des réseaux sans fil basés sur la technologie de transmission CDMA, est le support de multiples applications temps-réel telles la visiophonie, la télévision, etc. Nous étudions dans la suite un mécanisme de tarification qui permet dans un contexte de différenciation des services, de faire du contrôle dans une cellule d'un réseau sans fil CDMA. Nous considérons deux classes d'applications qui possèdent une sensibilité différente face à leur qualité de service respectives.

## 6.5 Tarification optimale multi-services

Nous considérons deux types d'applications avec différentes valuations de la qualité de service. Nous définissons ces deux types par «voix» et «données» en référence au chapitre 3. Nous supposons également que leur fonction d'utilité respective est décrite par l'équation (6.6). Les rapports puissance sur interférences plus bruit,  $SINR_i$ , pour les utilisateurs de type  $i \in \{v, d\}$  sont définis par les équations suivantes

$$\begin{aligned} SINR_v(N_v, N_d) &= \frac{P_v N_v}{\nu(N_v - 1)P_v + \nu N_d P_d + \chi^2}, \\ SINR_d(N_v, N_d) &= \frac{P_d N_d}{\nu N_v P_v + \nu(N_d - 1)P_d + \chi^2}. \end{aligned}$$

On remarque d'après l'expression des  $SINR$ , que l'utilité de chaque classe d'application dépend du nombre total d'utilisateurs dans la cellule. En effet, les interférences d'un utilisateur proviennent de tous les mobiles présents dans la cellule. Ainsi, le nombre d'utilisateurs de chaque type dans la cellule augmente tant que leur utilité respective est positive et que la contrainte de capacité n'est pas atteinte. Une augmentation du nombre d'utilisateurs d'un type diminue le rapport puissance sur interférences de l'autre type et donc diminue également son utilité. Nous cherchons alors l'équilibre  $(N_v^*, N_d^*)$  de ce système.

Nous supposons encore que pour chaque classe de trafic la capacité à la station de base est limitée. Ces contraintes s'écrivent à partir de l'équation (6.5)

$$N_v \leq \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_v} - N_d \frac{P_d}{P_v} + 1 \text{ (pour le trafic de type «voix»)}$$

et

$$N_v \leq \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_v} - N_d \frac{P_d}{P_v} + \frac{P_d}{P_v} \text{ (pour le trafic de type «données»)}.$$

Nous regroupons ces deux contraintes en l'inégalité suivante

$$N_v \leq \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_v} - N_d \frac{P_d}{P_v} + 1 + \frac{P_d - P_v}{P_v} \mathbb{1}_{P_d < P_v}. \quad (6.14)$$

Nous remarquons que lorsque le nombre d'utilisateurs à l'équilibre  $N_v^*$  et  $N_d^*$  annulent la contrainte de capacité (6.14), i.e.

$$N_v^* = \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_v} - N_d^* \frac{P_d}{P_v} + 1 + \frac{P_d - P_v}{P_v} \mathbb{1}_{P_d < P_v},$$

les débits moyen  $\lambda_v^c = \lambda_v(N_v^*, N_d^*)$  et  $\lambda_d^c = \lambda_d(N_v^*, N_d^*)$  ne dépendent pas de  $N_v^*$  et  $N_d^*$ . En effet, après quelques simplifications nous obtenons :

$$\lambda_v^c = \frac{lr}{\mathbb{E}D_v(\text{SINR}_v(N_v^*, N_d^*))} = \frac{rR_c}{n_v} \left( 1 - e^{-\omega \frac{n_v P_v}{\chi^2 / \eta + \nu(P_d - P_v) \mathbf{1}_{P_d < P_v}}} \right)^{lr},$$

et

$$\lambda_d^c = \frac{lr}{\mathbb{E}D_d(\text{SINR}_d(N_v^*, N_d^*))} = \frac{rR_c}{n_d} \left( 1 - e^{-\omega \frac{n_d P_d}{\chi^2 / \eta + \nu(P_v - P_d) \mathbf{1}_{P_v < P_d}}} \right)^{lr}.$$

Nous étudions le jeu non-coopératif entre les classes de trafic pour ce nouveau modèle avec différenciation de services en supposant tout d'abord que la demande pour chaque type d'applications excède toujours la capacité. Ensuite nous étudions, comme pour le cas mono-classe, le cas où les demandes pour chaque type sont représentées par des variables aléatoires discrètes positives et indépendantes.

### 6.5.1 Demande infinie

Les utilisateurs demandent à envoyer leurs paquets tant qu'ils y en ont un intérêt, c'est-à-dire que leur utilité est positive. De même, ils quittent le réseau si leur utilité est négative car leur satisfaction est alors inférieur au prix. Nous supposons, comme pour le cas mono-classe, que le système bloque l'entrée des utilisateurs si la contrainte de capacité définie par l'équation (6.14) est atteinte.

Le théorème suivant montre l'existence d'un unique équilibre de Nash  $(N_v^*, N_d^*)$  qui détermine le nombre d'utilisateurs pour chaque type de trafic à la stabilisation du système. Cet équilibre s'exprime en fonction des prix par paquets  $u_i$  et des puissances de réception  $P_i$  pour chaque classe  $i \in \{v, d\}$ .

**Théorème 17** *L'équilibre de Nash  $(N_v^*, N_d^*)$  dépend des puissances de réception  $P_v, P_d$  et des prix  $u_v, u_d$  de la façon suivante :*

1. Si  $U_v(1, 0) \leq 0 \Leftrightarrow u_v \geq V_v \left( \frac{n_v P_v}{\chi^2} \right)$  et  $U_d(0, 1) \leq 0 \Leftrightarrow u_d \geq V_d \left( \frac{n_d P_d}{\chi^2} \right)$ , aucun type d'utilisateurs ne possède un intérêt à rejoindre le réseau et donc

$$(N_v^*, N_d^*) = (0, 0).$$

2. Si  $U_v(1, 0) \leq 0 \Leftrightarrow u_v \geq V_v \left( \frac{n_v P_v}{\chi^2} \right)$  mais  $U_d(0, 1) > 0$ , seuls les utilisateurs de type «données» ont un intérêt à rejoindre le système et donc l'équilibre est  $(0, N_d^*)$  avec

$$N_d^* = \begin{cases} 1 - \frac{\chi^2}{\nu P_d} - \frac{n_d}{\nu V_d^{-1}(u_d)}, & \text{si } V_d\left(\frac{n_d P_d \eta}{\chi^2}\right) < u_d < V_d\left(\frac{n_d P_d}{\chi^2}\right), \\ 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_d}, & \text{sinon.} \end{cases} \quad (6.15)$$

3. De la même façon, si  $U_d(0,1) \leq 0 \Leftrightarrow u_d \geq V_d\left(\frac{n_d P_d}{\chi^2}\right)$  mais  $U_v(1,0) > 0$ , seuls les utilisateurs de type «voix» ont un intérêt à rejoindre le système et l'équilibre est  $(N_v^*, 0)$  avec

$$N_v^* = \begin{cases} 1 - \frac{\chi^2}{\nu P_v} - \frac{n_v}{\nu V_v^{-1}(u_v)}, & \text{si } V_v\left(\frac{n_v P_v \eta}{\chi^2}\right) < u_v < V_v\left(\frac{n_v P_v}{\chi^2}\right), \\ 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v}, & \text{sinon.} \end{cases} \quad (6.16)$$

4. Enfin, si  $U_v(1,0) > 0$  et  $U_d(0,1) > 0$ , nous avons 3 possibilités.

- (a) Si  $\frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v} \leq \min\left(\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d, \frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v\right)$ , la contrainte de capacité (6.14) est atteinte alors que les utilisateurs ont un intérêt à rejoindre le réseau. Alors le système choisit d'accepter le type d'utilisateurs qui maximise le revenu de la station de base. Ainsi, il choisit seulement les utilisateurs de type «données» si  $u_d > \frac{P_d \lambda_v^c}{P_v \chi_d^c} u_v$  ou les utilisateurs de type «voix» sinon.

L'équilibre est donc  $(0, N_d^*)$  avec  $N_d^* = 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_d}$  si  $u_d > \frac{P_d \lambda_v^c}{P_v \chi_d^c} u_v$ , et  $(N_v^*, 0)$  avec  $N_v^* = 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v}$  sinon.

- (b) Si  $\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d < \min\left(\frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v}, \frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v\right)$ , l'équilibre est  $(N_v^*, 0)$  avec  $N_v^*$  défini par l'équation (6.16).

- (c) sinon  $\frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v < \min\left(\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d, \frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v}\right)$  et l'équilibre est  $(0, N_d^*)$  avec  $N_d^*$  défini par l'équation (6.15).

**Preuve** La preuve consiste à étudier les différents cas suivants les prix et les puissances.

1. Si  $U_v(1,0) \leq 0 \Leftrightarrow u_v \geq V_v\left(\frac{n_v P_v}{\chi^2}\right)$  et  $U_d(0,1) \leq 0 \Leftrightarrow u_d \geq V_d\left(\frac{n_d P_d}{\chi^2}\right)$ , aucun utilisateur n'a intérêt à rejoindre le système car son utilité est négative. Comme les fonctions d'utilité  $U_v$  et  $U_d$  sont décroissantes en  $N_v$  et  $N_d$ , ce résultat est valable pour tout couple  $(N_v, N_d)$  d'utilisateurs présents dans le système. Ainsi, pour cette configuration de prix et de puissance, l'équilibre est

$$(N_v^*, N_d^*) = (0, 0).$$

2. Si  $U_v(1,0) \leq 0 \Leftrightarrow u_v \geq V_v\left(\frac{n_v P_v}{\chi^2}\right)$  mais  $U_d(0,1) > 0$ , seuls les utilisateurs de type «données» requièrent le service car  $U_v(N_v, N_d) \leq 0 \forall (N_v, N_d)$ . Par conséquent, nous nous ramenons au cas d'une seule classe de trafic étudiée dans le théorème 14. L'équilibre est donc  $(0, N_d^*)$  avec

$$N_d^* = \begin{cases} 1 - \frac{\chi^2}{\nu P_d} - \frac{n_d}{\nu V_d^{-1}(u_d)}, & \text{si } V_d\left(\frac{n_d P_d \eta}{\chi^2}\right) < u_d < V_d\left(\frac{n_d P_d}{\chi^2}\right), \\ 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_d}, & \text{sinon.} \end{cases} \quad (6.17)$$

3. De façon symétrique, si  $U_d(0,1) \leq 0 \Leftrightarrow u_d \geq V_d\left(\frac{n_d P_d}{\chi^2}\right)$  mais  $U_v(1,0) > 0$ , nous utilisons le théorème 14 avec une unique classe de trafic de type «voix» et l'équilibre est  $(N_v^*, 0)$  avec

$$N_v^* = \begin{cases} 1 - \frac{\chi^2}{\nu P_v} - \frac{n_v}{\nu V_v^{-1}(u_v)}, & \text{si } V_v\left(\frac{n_v P_v \eta}{\chi^2}\right) < u_v < V_v\left(\frac{n_v P_v}{\chi^2}\right), \\ 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v}, & \text{sinon.} \end{cases} \quad (6.18)$$

4. Enfin, si  $U_v(1,0) > 0$  et  $U_d(0,1) > 0$ , les deux type d'applications ont un intérêt à envoyer leurs paquets à la station de base.

La contrainte de capacité définie par l'équation (6.14) s'exprime par la combinaison linéaire suivante :

$$N_v P_v + N_d P_d = \frac{1-\eta}{\eta} \frac{\chi^2}{\nu} + P_v + (P_d - P_v) \mathbb{1}_{P_d < P_v}.$$

On note également que les équations  $U_d(N_v, N_d) = 0$  et  $U_v(N_v, N_d) = 0$  s'écrivent sous la même forme

$$U_d(N_v, N_d) = 0 \Leftrightarrow N_v P_v + N_d P_d = \frac{n_d P_d}{\nu V_d^{-1}(u_d)} - \frac{\chi^2}{\nu} + P_d. \quad (6.19)$$

et

$$U_v(N_v, N_d) = 0 \Leftrightarrow N_v P_v + N_d P_d = \frac{n_v P_v}{\nu V_v^{-1}(u_v)} - \frac{\chi^2}{\nu} + P_v. \quad (6.20)$$

Les utilisateurs entrent jusqu'à ce qu'une des précédentes conditions soit vérifiée. Ces équations définissent trois droites, plus exactement segments, parallèles sur le domaine  $N_v \geq 0, N_d \geq 0$ . Les différentes positions des ces droites les unes par rapport aux autres sont représentées sur la figure 6.5. Nous avons différents équilibres suivant quelle droite est la plus "basse", c'est-à-dire la plus proche du point  $(0, 0)$ .

- (a) Tout d'abord, si  $\frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v) \mathbb{1}_{P_d < P_v} \leq \min\left(\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d, \frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v\right)$ , alors la contrainte de capacité donnée par l'équation (6.14) est la plus basse (cas (a) de la figure 6.5). Donc le système accepte  $N_v + N_d$  utilisateurs tel que la contrainte soit active, c'est-à-dire :

$$N_v P_v + N_d P_d = \frac{1-\eta}{\eta} \frac{\chi^2}{\nu} + P_v + (P_d - P_v) \mathbb{1}_{P_d < P_v}. \quad (6.21)$$

Il existe donc une infinité d'équilibres et nous choisissons l'équilibre qui optimise le revenu de la station de base. Celui-ci s'exprime en fonction du nombre d'utilisateurs de chaque type par

$$N_v \lambda_v u_v + N_d \lambda_d u_d = N_d (u_d \lambda_d^c - u_v \lambda_v^c \frac{P_d}{P_v}) + u_v \lambda_v^c \left( \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v} + 1 + \frac{P_d - P_v}{P_v} \mathbb{1}_{P_d < P_v} \right)$$

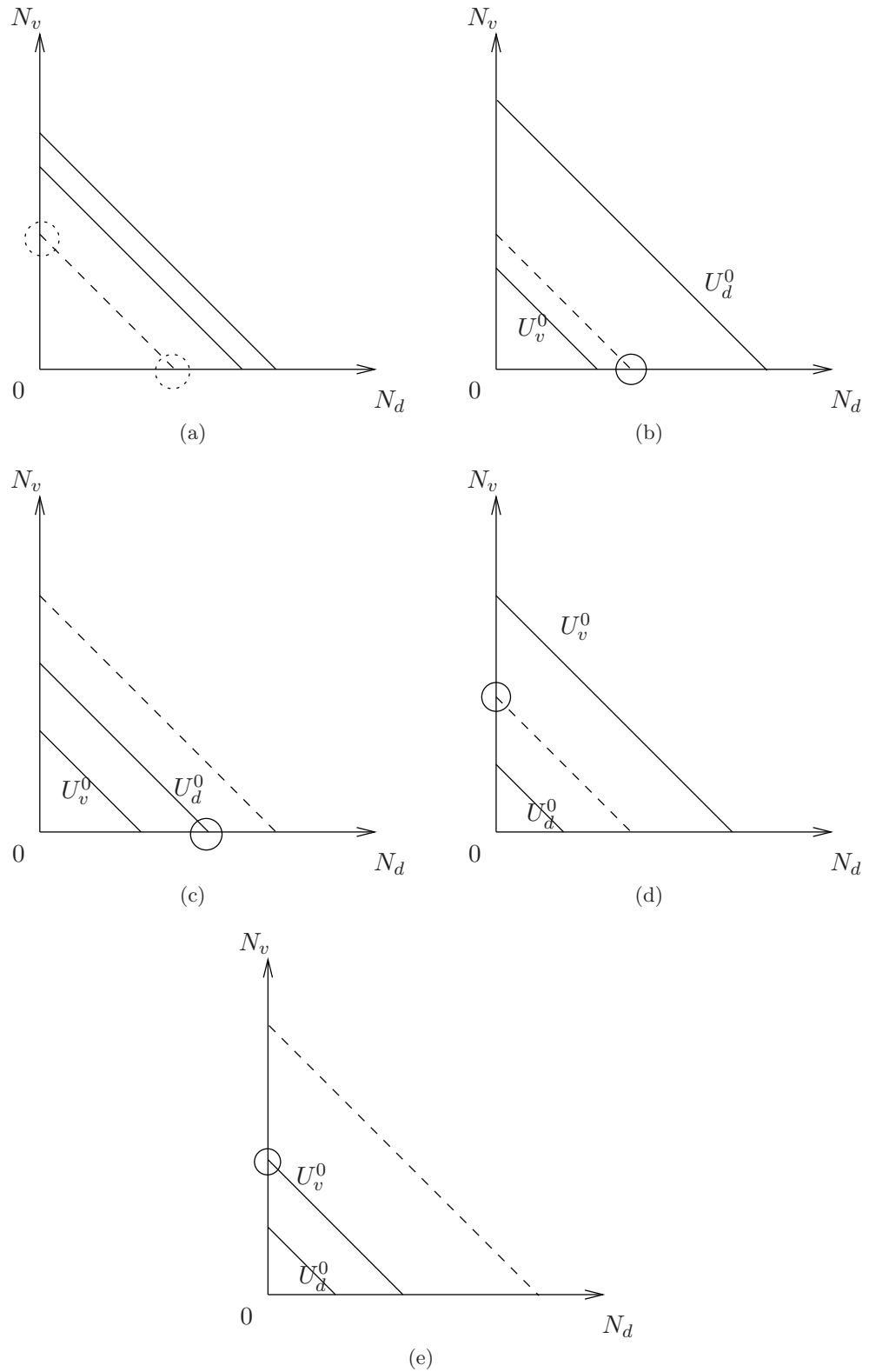


FIG. 6.5 – Équilibre en fonction des prix et des puissances de réception lorsque la demande excède toujours la capacité. La droite en pointillé représente la contrainte de capacité. La droite  $U_v^0$  représente l'équation  $U_v(N_v, N_d) = 0$  et  $U_d^0$  représente  $U_d(N_v, N_d) = 0$ . Les cercles dénotent les équilibres de Nash du système.

avec

$$\lambda_v^c = \frac{rR_c}{n_v} \left(1 - e^{-\omega \frac{n_v P_v}{\chi^2 / \eta + \nu(P_d - P_v) \mathbb{1}_{P_d < P_v}}} \right) l r$$

et

$$\lambda_d^c = \frac{rR_c}{n_d} \left(1 - e^{-\omega \frac{n_d P_d}{\chi^2 / \eta + \nu(P_v - P_d) \mathbb{1}_{P_v < P_d}}} \right) l r.$$

Le revenu de la station de base est alors linéaire en  $N_d$  et le système choisit l'équilibre qui maximise son revenu. Cela dépend du signe de  $u_d \lambda_d^c - u_v \lambda_v^c \frac{P_d}{P_v}$ . Ainsi, si  $u_d \lambda_d^c - u_v \lambda_v^c \frac{P_d}{P_v} < 0$  le système choisit l'équilibre  $(N_v^*, 0)$  avec

$$N_v^* = 1 + \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_v},$$

car le revenu est décroissant en  $N_d$ . Sinon, l'équilibre est  $(0, N_d^*)$  avec

$$N_d^* = 1 + \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_d}.$$

- (b) Si  $\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d < \min\left(\frac{\chi^2}{\eta \nu} + P_v + (P_d - P_v) \mathbb{1}_{P_d < P_v}, \frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v\right)$ , la droite  $U_d(N_v, N_d) = 0$  est la plus basse (cas (d) et (e) de la figure 6.5). Alors le nombre d'utilisateurs de chaque type augmente jusqu'à  $(N_v, N_d)$  tel que  $U_d(N_v, N_d) = 0$ . Or, nous avons  $U_v(N_v, N_d) > 0$  et donc le nombre d'utilisateurs de type «voix» augmente, ce qui rend l'utilité de type «données» négative. Ainsi, des utilisateurs de type «données» quittent le système afin que  $U_d$  soit égale à zéro. Ainsi, la dynamique du nombre d'utilisateurs est de glisser le long de la droite  $U_d = 0$  jusqu'à  $N_d = 0$ . Finalement, le nombre d'utilisateurs  $N_v$  de type «voix» augmente (le long de l'axe des ordonnées sur la figure 6.5) jusqu'à atteindre la limite de capacité (cas (d) de la figure 6.5) ou la droite  $U_v(N_v, 0) = 0$  (cas (e) de la figure 6.5).

L'équilibre est donc  $(N_v^*, 0)$  avec

$$N_v^* = \begin{cases} 1 - \frac{\chi^2}{\nu P_v} - \frac{n_v}{\nu V_v^{-1}(u_v)}, & \text{si } V_v\left(\frac{n_v P_v \eta}{\chi^2}\right) < u_v < V_v\left(\frac{n_v P_v}{\chi^2}\right), \\ 1 + \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_v}, & \text{sinon.} \end{cases} \quad (6.22)$$

- (c) Sinon  $\frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v < \min\left(\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d, \frac{\chi^2}{\eta \nu} + P_v + (P_d - P_v) \mathbb{1}_{P_d < P_v}\right)$ , le résultat s'obtient de la même façon que le cas précédent en inversant le type «voix» et le type «données» (cas (b) et (c) de la figure 6.5). On obtient donc l'équilibre  $(0, N_d^*)$  avec

$$N_d^* = \begin{cases} 1 - \frac{\chi^2}{\nu P_d} - \frac{n_d}{\nu V_d^{-1}(u_d)}, & \text{si } V_d\left(\frac{n_d P_d \eta}{\chi^2}\right) < u_d < V_d\left(\frac{n_d P_d}{\chi^2}\right), \\ 1 + \frac{1 - \eta}{\eta} \frac{\chi^2}{\nu P_d}, & \text{sinon.} \end{cases} \quad (6.23)$$

■



Suivant les prix et les puissances de réception, l'équilibre de Nash  $(N_v^*, N_d^*)$  du système est donc différent mais on remarque que seule une classe d'application est en service, i.e.  $N_v^* = 0$  ou  $N_d^* = 0$ . Nous pouvons alors déterminer aisément les prix et puissances qui maximisent le revenu du système. En effet, le système à l'équilibre est étudié dans le théorème 15 pour une unique classe de service. Nous avons le théorème suivant qui donne le résultat de cette optimisation.

**Théorème 18** *Le revenu maximum de la station de base est*

$$R^* = \max(R_v^*, R_d^*),$$

où  $R_v^*$  (resp.  $R_d^*$ ) est le revenu maximum avec une unique classe d'applications de type «voix» (resp. «données») obtenu dans le théorème 15. Soient  $i \in \{v, d\}$  tel que  $R_i^* = R^*$  et  $j \in \{v, d\}$  tel que  $j \neq i$ . Le prix par paquet pour la classe  $i$  est  $u_i^* = V_i(X^*)$  et la puissance de réception optimale  $P_i^* = \frac{X^* \chi^2}{n_i \eta}$  avec  $X^*$  solution de l'équation (6.9) adaptée à la classe  $i$ .

Les valeurs du prix optimal  $u_j^*$  et de la puissance optimale  $P_j^*$  pour l'autre classe  $j$  de trafic sont déterminés de sorte que l'équilibre qui en résulte vérifie  $N_j^* = 0$ . Ainsi,  $\forall P_j^* > 0$ , il faut que :

$$u_j^* \in \left[ 0, \min \left( \frac{\lambda_i^c P_j^*}{\lambda_j^c P_i^*} u_i^*, V_j \left( \frac{n_j P_j^*}{\chi^2} \eta \right) \right) \right] \cup \left[ V_j \left( \frac{n_j P_j^*}{\chi^2} \eta \right), +\infty \right[$$

**Preuve** Le théorème 17 montre qu'à l'équilibre, seul un type d'application est présent ou envoie ses paquets. On définit par  $i \in \{v, d\}$  le type d'application tel que  $R_i^* = \max(R_v^*, R_d^*)$  et donc le système définit les prix et les puissances de sorte que seul les utilisateurs de ce type rejoignent le système à l'équilibre. Le théorème 15 donne le prix par paquet et la puissance de réception optimaux pour la classe  $i$  :

$$u_i^* = V_i(X^*) \quad \text{et} \quad P_i^* = \frac{X^* \chi^2}{n_i \eta},$$

avec  $X^*$  défini par :

$$X^* = \arg \max_{X \in \mathcal{X}} \left( \frac{r R_c}{n_i} (1 - e^{-\omega X})^{lr} V_i(X) + \frac{1 - \eta}{\nu} r R_c \frac{(1 - e^{-\omega X})^{lr} V_i(X)}{X} \right).$$

Enfin, le prix optimal  $u_j^*$  et la puissance optimale  $P_j^*$  avec  $j \in \{v, d\}$ ,  $j \neq i$ , doivent être choisis de sorte que seuls les utilisateurs de classe  $i$  soient présents dans le système à l'équilibre.

Supposons que le revenu optimal de la station de base correspond au cas où seul le trafic de type «voix» est présent dans le système à l'équilibre. Alors, d'après le théorème 15, nous avons

$$u_v^* = V_v(X^*) \quad \text{et} \quad P_v^* = \frac{X^* \chi^2}{n_v \eta}.$$

Le système doit maintenant déterminer le prix  $u_d^*$  et la puissance de réception  $P_d^*$  pour le trafic de type «données» qui conduisent, avec  $u_v^*$  et  $P_v^*$ , à l'équilibre  $(N_v^*, 0)$ . Nous déterminons l'équilibre du système en observant la position des 3 droites  $U_v^0$ ,  $U_d^0$  et la contrainte de capacité (voir figure 6.5). Il suffit alors de savoir laquelle est la plus "basse" en comparant leur ordonnées à l'origine. D'après l'équation (6.20), l'abscisse à l'origine de la droite  $U_v = 0$  est :

$$U_v^0 = 1 + (1 - \eta) \frac{n_v}{\nu X^*}.$$

D'après l'équation (6.14), l'abscisse à l'origine pour la contrainte de capacité correspond à :

$$C^0 = 1 + (1 - \eta) \frac{n_v}{\nu X^*} + \frac{P_d^* - P_v^*}{P_v^*} \mathbb{1}_{P_d < P_v^*}.$$

Nous avons  $U_v^0 \geq C^0$ , ce qui implique que la configuration des 3 droites correspond aux cas (a) ou (d) de la figure 6.5.

- Si  $U_d^0 < C^0 \Leftrightarrow u_d > V_d(\frac{n_d P_d^*}{\chi^2} \eta)$  l'équilibre est alors  $(N_v^*, 0)$  car nous sommes dans la configuration (d).
- Sinon  $u_d \leq V_d(\frac{n_d P_d^*}{\chi^2} \eta)$  et alors l'équilibre est  $(N_v^*, 0)$  si et seulement si  $u_d < \frac{\lambda_v^c P_d^*}{\lambda_d^c P_v^*} u_v^*$ , d'après la preuve du théorème 17.

Donc, si  $R^* = R_i^*$ , l'ensemble des prix optimaux et des puissances optimales est :

$$u_i^* = V_i(X^*), \quad P_i^* = \frac{X^* \chi^2}{n_i \eta},$$

et

$$P_j^* > 0, \quad u_j^* \in \left[ 0, \min \left( \frac{\lambda_i^c P_j^*}{\lambda_j^c P_i^*} u_i^*, V_j \left( \frac{n_j P_j^*}{\chi^2} \eta \right) \right) \right] \cup \left[ V_j \left( \frac{n_j P_j^*}{\chi^2} \eta \right), +\infty [ . \quad \blacksquare$$

*Exemple* Nous illustrons ce résultat en considérant l'application numérique avec la fonction d'utilité donnée par l'équation (6.7). Nous reprenons les données numériques de l'exemple précédent pour le cas avec un seul type d'application, lesquelles correspondent ici au type «données». Nous considérons pour le trafic de type «voix», le facteur d'étalement  $n_v = 8$  et la sensibilité au délai  $\alpha_v = 2$ . Ainsi, ce type de trafic est supposé plus sensible à un délai important que le trafic de type «données» car  $\alpha_v > \alpha_d$ . Nous obtenons que seul le trafic de type «voix» est présent à l'équilibre, le prix par paquet optimal est  $u_v^* = 0.632$ , la puissance de réception  $P_v^* = 2.737$  et le revenu maximum du fournisseur est alors  $R^* = 574.603$ . De plus, en considérant par exemple une puissance de réception  $P_d^* = 3$ , l'intervalle du prix  $u_d^*$  est  $[0, +\infty[$  et donc, pour n'importe quel prix pour le trafic de type «données», l'équilibre est toujours composé uniquement du trafic de type «voix».

Nous avons montré dans cette partie que lorsque la demande pour chaque type d'application excède toujours la capacité, le revenu de la station de base est maximisé avec une unique classe de trafic à l'équilibre. Dans un contexte plus réaliste, la demande

varie dans le temps et peut donc être inférieure à la capacité à certains instants. Nous modélisons cette variabilité de la demande par une variable aléatoire positive discrète, comme nous l'avons fait dans la partie avec une unique classe de service. Nous supposons que la demande pour chaque type est une variable aléatoire et qu'elles sont indépendantes l'une de l'autre. Nous étudions ce nouveau modèle dans la section suivante et nous montrons alors que, contrairement au cas où la demande excède toujours la capacité, le revenu peut-être optimisé avec les deux types d'applications en service.

### 6.5.2 Demande aléatoire

Nous supposons que pour tout  $i \in \{v, d\}$ , la demande ou nombre d'utilisateurs de type  $i$  voulant envoyer leurs paquets à la station de base, est donnée par une variable aléatoire positive discrète  $\Delta_i$ . De plus, les variables  $\Delta_v$  et  $\Delta_d$  sont supposées indépendantes. Suivant les réalisations  $\delta_i$  des variables  $\Delta_i$ , il existe un équilibre de Nash pour le système définissant le nombre d'utilisateurs  $\kappa_v^*$  de type «voix» et  $\kappa_d^*$  de type «données».

**Théorème 19** *Soient  $\delta_v$  et  $\delta_d$  deux réalisations respectives des demandes aléatoires du trafic de type «voix» et «données». Il existe un équilibre (de Nash)  $(\kappa_v^*, \kappa_d^*)$  déterminant le nombre d'utilisateurs de type «voix» et «données» présents dans le système. Cet équilibre dépend des puissances de réception  $P_v, P_d$ , des prix  $u_v, u_d$ , et des réalisations  $\delta_v, \delta_d$  des demandes de chaque type, de la façon suivante :*

1. si  $(\delta_v = 0$  ou  $U_v(1, 0) \leq 0 \Leftrightarrow u_v \geq V_v \left( \frac{n_v P_v}{\chi^2} \right))$  et  $(\delta_d = 0$  ou  $U_d(0, 1) \leq 0 \Leftrightarrow u_d \geq V_d \left( \frac{n_d P_d}{\chi^2} \right))$ , aucun type d'utilisateurs ont intérêt à envoyer leurs paquets et donc l'équilibre est

$$(\kappa_v^*, \kappa_d^*) = (0, 0).$$

2. Si  $\delta_v = 0$  ou  $U_v(1, 0) \leq 0 \Leftrightarrow u_v \geq V_v \left( \frac{n_v P_v}{\chi^2} \right)$  mais  $\delta_d > 0$  et  $U_d(0, 1) > 0$ , alors seul le trafic de type «données» est présent (à l'équilibre) et l'équilibre est  $(0, \kappa_d^*)$  avec

$$\kappa_d^* = \min(\delta_d, N_d^*), \quad (6.24)$$

où  $N_d^*$  est défini dans l'équation (6.23).

3. De façon symétrique, si  $\delta_d = 0$  ou  $U_d(0, 1) \leq 0 \Leftrightarrow u_d \geq V_d \left( \frac{n_d P_d}{\chi^2} \right)$  mais  $\delta_v > 0$  et  $U_v(1, 0) > 0$ , alors seul le trafic de type «voix» est présent et l'équilibre est alors  $(\kappa_v^*, 0)$  avec

$$\kappa_v^* = \min(\delta_v, N_v^*), \quad (6.25)$$

où  $N_v^*$  est défini dans l'équation (6.22).

4. Si la demande totale est inférieure à la capacité, c'est-à-dire  $\delta_v < \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v} - \delta_d \frac{P_d}{P_v} + 1 + \frac{P_d - P_v}{P_v} \mathbb{1}_{P_d < P_v}$  et  $U_i(\delta_v, \delta_d) > 0$  pour  $i = \{v, d\}$ , alors l'équilibre est

$$(\kappa_v^*, \kappa_d^*) = (\delta_v, \delta_d).$$

5. Sinon, nous avons les trois cas de figures suivants :

- (a) si  $\frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v} \leq \min\left(\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d, \frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v\right)$ , la contrainte de capacité définie par la relation (6.14) et les contraintes de demande  $\delta_v$  et  $\delta_d$  sont atteintes alors que les utilisateurs de chaque type ont intérêt à rejoindre le réseau. L'équilibre est donc  $(\kappa_v^*, \kappa_d^*)$  avec
- $\kappa_d^* = \min(\delta_d, \frac{P_v}{P_d} + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_d} + \frac{(P_d - P_v)}{P_d} \mathbb{1}_{P_d < P_v})$  et
  - $\kappa_v^* = \frac{\frac{1-\eta}{\eta} \frac{\chi^2}{\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v} - \kappa_d^* P_d}{P_v}$  si  $u_d > \frac{P_d \lambda_v^c}{P_v \lambda_d^c} u_v$ ,
  - $\kappa_v^* = \min(\delta_v, 1 + \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v} + \frac{(P_d - P_v)}{P_v} \mathbb{1}_{P_d < P_v})$  et
  - $\kappa_d^* = \frac{\frac{1-\eta}{\eta} \frac{\chi^2}{\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v} - \kappa_v^* P_v}{P_d}$  sinon.
- (b) Si  $\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d < \min\left(\frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v}, \frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v\right)$  l'équilibre est  $(\kappa_v^*, \max(0, \frac{n_d}{\nu V_d^{-1}(u_d)} - \delta_v \frac{P_v}{P_d} - \frac{\chi^2}{\nu P_d} + 1))$  avec  $\kappa_v^*$  défini dans (6.25).
- (c) Sinon  $\frac{n_v P_v}{\nu V_v^{-1}(u_v)} + P_v < \min\left(\frac{n_d P_d}{\nu V_d^{-1}(u_d)} + P_d, \frac{\chi^2}{\eta\nu} + P_v + (P_d - P_v)\mathbb{1}_{P_d < P_v}\right)$ , l'équilibre est  $(\max(0, \frac{n_v}{\nu V_v^{-1}(u_v)} - \delta_d \frac{P_d}{P_v} - \frac{\chi^2}{\nu P_v} + 1), \kappa_d^*)$  avec  $\kappa_d^*$  défini dans (6.24).

**Preuve** L'équilibre est déterminé en fonction de la position des droites  $U_v = 0, U_d = 0$  et de la contrainte de capacité, les unes par rapport aux autres. Prenons l'exemple de configuration des droites représentées sur le cas (a) de la figure 6.5. Les autres configurations (b), (c), (d) et (e) s'analysent de la même façon. Suivant les demandes  $\delta_v$  et  $\delta_d$ , nous avons différents équilibres représentés sur la figure 6.6.

Lorsque la demande totale  $\delta_v + \delta_d$  est inférieure à la capacité, et que chaque type de trafic possède un intérêt à envoyer ses paquets, alors l'équilibre correspond à la demande de chacun  $N_v^* = \delta_v$  et  $N_d^* = \delta_d$  (cas (a) de la figure 6.6). Cette configuration correspond au quatrième point du théorème précédent car d'après l'équation de la contrainte de capacité (6.14), la demande totale plus petite que la capacité est équivalent à :

$$\delta_v < \frac{1-\eta}{\eta} \frac{\chi^2}{\nu P_v} - \delta_d \frac{P_d}{P_v} + 1 + \frac{P_d - P_v}{P_v} \mathbb{1}_{P_d < P_v}.$$

Il faut noter que les utilisateurs entrent dans le système tant qu'ils y ont un intérêt (utilité positive), que la capacité n'est pas atteinte (contrainte de capacité inactive) et que leur demande n'est pas satisfaite (nombre d'utilisateurs présents plus petit que le nombre qui demande le service).

Ainsi pour le cas (b) (resp. (c)), la demande pour les utilisateurs de type «voix» (resp. «données») est plus importante que pour le cas (a) et ainsi la demande totale est supérieure à la capacité. Pour le cas (d), la demande de chaque type est supérieure à la capacité. Dans tous les cas, l'équilibre annule la contrainte de capacité. Le système possède alors une infinité d'équilibre et décide de choisir la classe de trafic qui optimise son revenu (les cercles en pointillé).

Nous avons donc déterminé quels sont les équilibres pour la configuration (a) des droites de la figure 6.5. Il reste à étudier les 4 différents cas pour les 4 autres configurations de la figure 6.5, c'est-à-dire 16 possibilités au total.

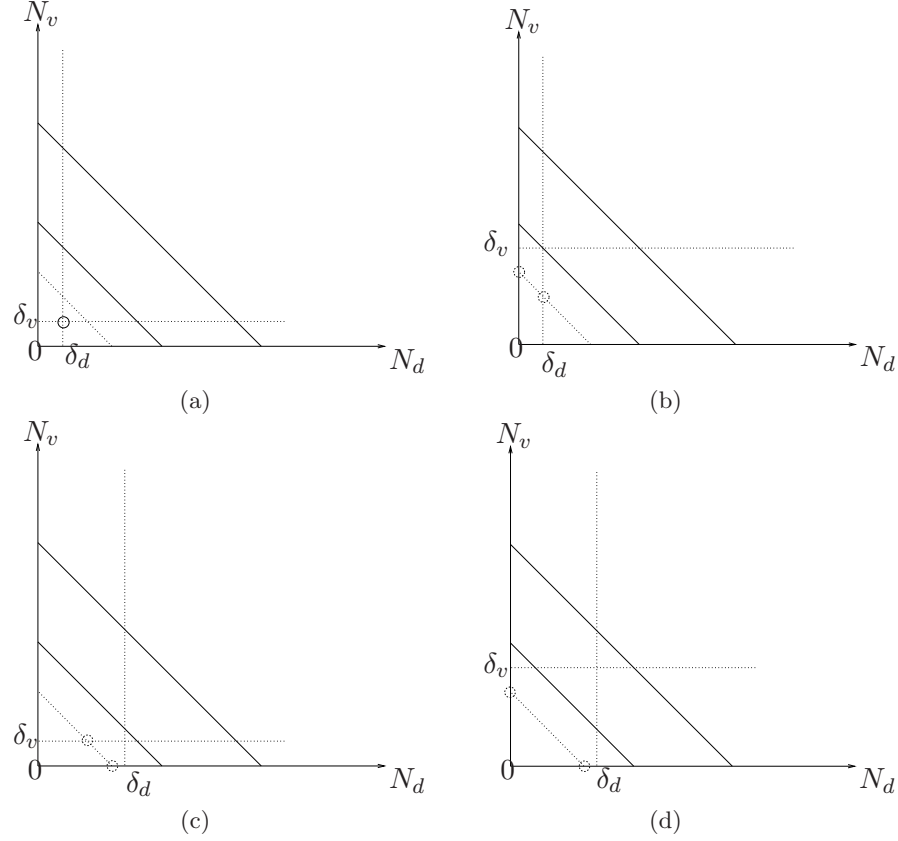


FIG. 6.6 – Équilibre en fonction des prix et des puissances de réception lorsque la demande est aléatoire. La droite en pointillée représente la contrainte de capacité. La droite  $U_v^0$  représente l'équation  $U_v(N_v, N_d) = 0$  et  $U_d^0$  représente  $U_d(N_v, N_d) = 0$ . Les cercles dénotent les équilibres de Nash du système.

Nous ne décrivons pas ici, pour des raisons de simplicité, les 16 autres configurations car le raisonnement est identique au précédent. Les prix, les puissance et les demandes induisent des configurations différentes des droites mais, ensuite la dynamique du jeu qui conduit à l'équilibre de Nash entre les classes est la même. ■

Une fois déterminé l'équilibre du système, la station de base s'intéresse à optimiser son revenu moyen en fonction des prix par paquet et des puissances de réception pour chaque type d'applications. Le revenu moyen de la station de base s'exprime par :

$$\begin{aligned} \bar{R}(u_v, P_v, u_d, P_d) &= \sum_{\delta_v=0}^{\infty} \sum_{\delta_d=0}^{\infty} (u_v \kappa_v^*(u_v, P_v, \delta_v, u_d, P_d, \delta_d) \lambda_v(u_v, P_v, \delta_v, u_d, P_d, \delta_d) \\ &\quad + u_d \kappa_d^*(u_v, P_v, \delta_v, u_d, P_d, \delta_d) \lambda_d(u_v, P_v, \delta_v, u_d, P_d, \delta_d)) \mathbb{P}(\Delta_v = \delta_v, \Delta_d = \delta_d), \end{aligned}$$

où  $\forall i \in \{v, d\}$ ,  $\kappa_i^*(u_v, P_v, \delta_v, u_d, P_d, \delta_d)$  (resp.  $\lambda_i(u_v, P_v, \delta_v, u_d, P_d, \delta_d)$ ) est le nombre d'utilisateurs (resp. le débit moyen) de type  $i$  à l'équilibre.

Une étude analytique de cette optimisation semble infaisable à cause de l'expression complexe de l'équilibre qui pose des problèmes de dérivation. Nous utilisons l'algorithme 1 d'optimisation stochastique de type recuit simulé défini dans le chapitre 2 car nous avons remarqué que les algorithmes déterministes convergent vers des maxima locaux multiples.

*Exemple* Nous considérons l'ensemble des paramètres de l'application numérique précédente et supposons de plus que la demande suit une loi de Poisson de taux 2 pour le trafic de type «voix» et 3 pour le type «données». Alors, nous obtenons un revenu moyen maximum  $R^* = 504.836$  pour des prix optimaux  $u_v^* = 0.638$ ,  $u_d^* = 0.484$  et des puissances optimales  $P_v^* = 2.635$  et  $P_d^* = 2.177$ . Nous représentons sur la figure 6.7 le revenu moyen en fonction du prix par paquet et de la puissance de réception pour les utilisateurs de type «voix» avec les prix et puissances optimaux pour le trafic de type «données».

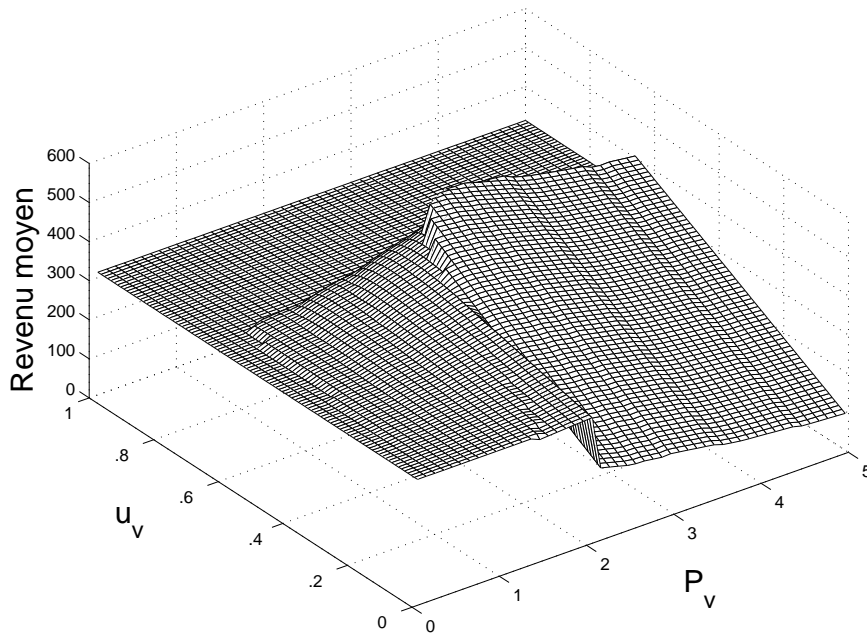


FIG. 6.7 – Revenu moyen de la station de base en fonction des prix  $u_v$  et puissances  $P_v$  pour le trafic de type «voix» avec le prix optimal  $u_d^*$  et la puissance optimale  $P_d^*$  pour le trafic de type «données».

Nous représentons également le revenu moyen en fonction cette fois-ci des puissances sur le figure 6.8. Les illustrations sont bien en accord avec l'optimum déterminé par l'algorithme d'optimisation.

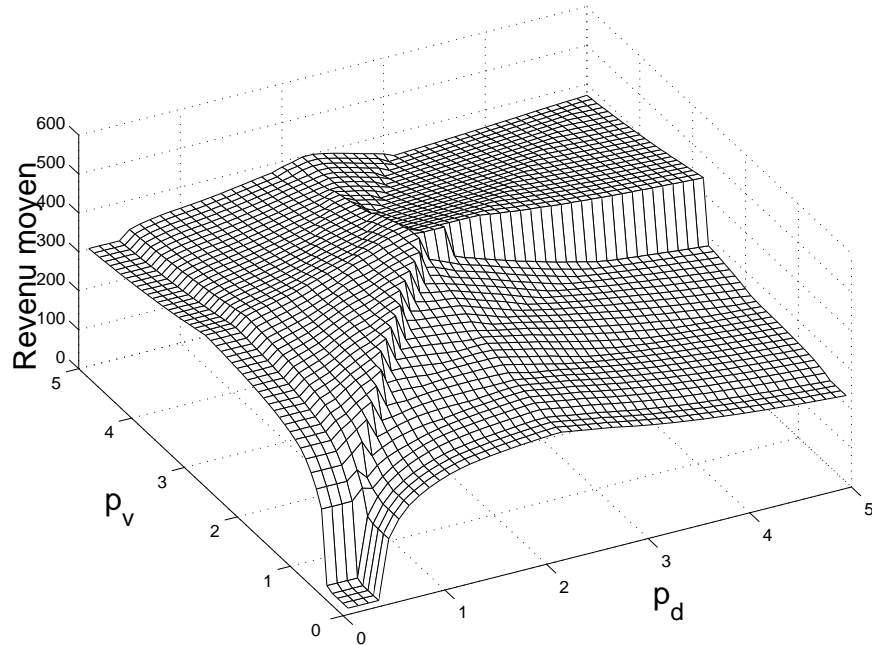


FIG. 6.8 – Revenu moyen de la station de base en fonction des puissances  $P_v$  et  $P_d$  en utilisant les prix optimaux  $u_v^*$  et  $u_d^*$ .

## 6.6 Conclusion

Les réseaux sans fil, et particulièrement ceux de troisième génération sont fortement sujets au problème de congestion dû à la limitation de la ressource radio et aux nouvelles applications multimédia telles la visiophonie ou la télévision qui requièrent des débits importants. Nous avons montré dans cette partie comment un mécanisme de tarification à l'usage permet tout d'abord de contrôler la demande et donc le trafic dans ce type de réseau, puis d'engendrer du profit pour le fournisseur. Nous nous sommes intéressés au trafic généré sur la voie montante, des mobiles vers la station de base dans une cellule d'un réseau sans fil qui utilise la technologie CDMA.

Nous proposons une tarification statique à l'usage pour une cellule CDMA avec plusieurs classes de service où le contrôle de puissance est supposé parfait. Les modèles de tarification proposés dans la littérature pour les réseaux cellulaires CDMA basent leurs prix sur la puissance d'émission des mobiles afin de contrôler le niveau d'interférences. Le prix est alors dynamique, évoluant suivant la puissance d'émission du mobile qui dépend de la distance du mobile à la station de base. Nous proposons dans ce chapitre une tarification statique qui devrait être mieux appréciée par les utilisateurs qu'une tarification dynamique, car nos prix sont fixés [41]. La station de base adapte les prix et les puissances de réception afin de maximiser son revenu. Nous observons que cette tarification à l'usage engendre un jeu non-coopératif entre les différents types d'utili-

sateurs ou applications. Nos contributions porte d'abord sur l'analyse de ce jeu. Nous montrons qu'il possède un unique équilibre de Nash et ainsi, le nombre d'utilisateurs de chaque type se stabilise à cet état d'équilibre en fonction des prix et puissances déterminés par la station de base. Nous déterminons ensuite les paramètres, prix et puissance, pour chaque classe de trafic qui optimisent le revenu de la station de base en considérant l'équilibre de Nash sous-jacent. Nous considérons de plus deux scénarii de demande. Tout d'abord, le cas où le nombre d'utilisateurs est toujours supérieur à la capacité et est donc infini. Ensuite, le cas plus réaliste où ce nombre est modélisé par une variable aléatoire, ce qui permet de prendre en compte la variabilité de la demande au cours du temps.



## Chapitre 7

# Allocation optimale avec choix aléatoire de la classe de service

Dans le chapitre précédent, afin de modéliser la variabilité de la demande, nous avons modélisé le nombre d'utilisateurs voulant rejoindre le système par une variable aléatoire. Dans ce dernier chapitre, les utilisateurs ont un comportement (utilité) non-déterministe contrairement aux études précédentes et le système optimise son revenu moyen suivant les probabilités de choix des utilisateurs. Ces probabilités dépendent explicitement des utilités pour chaque classe de service et donc les demandes pour chaque classe sont dépendantes. Le fournisseur possède une quantité finie  $C$  de ressource (bande passante) et son objectif est de déterminer une allocation optimale de cette ressource afin de maximiser son revenu. La contribution de ce travail est de considérer un modèle de choix discret probabiliste qui dépend du prix et de l'allocation de chaque segment. Ce travail est le fruit d'une collaboration avec les chercheurs de IBM T.J. Watson Research Center et a été publié dans [39].

### 7.1 Le *Yield Management* et la tarification des réseaux

Le «*yield management*» (désigné également sous le terme «*revenue management*») est une technique qui permet de calculer les prix optimisant le profit généré par la vente d'un produit ou d'un service, sur la base d'une modélisation et d'une prévision du comportement de la demande par micro segment du marché. Cette méthode de calcul des prix a révolutionné le transport aérien au début des années 80 [82]. Elle permet de résoudre de manière optimale le problème de la confrontation de l'offre et de la demande, grâce à une tarification différenciée et au contrôle de la quantité (allocation) d'un produit mis en vente dans chaque classe tarifaire appelée segment. Les conséquences positives de l'utilisation de ce concept économique sont ressenties par l'ensemble des intervenants : côté fournisseur, hausse du profit ; côté utilisateur, possibilité de baisse des prix sans impact sur la qualité de service. Depuis une dizaine d'années, cette technique de marketing pénètre d'autres secteurs d'activités comme l'hôtellerie.

## 7.2 Modèle de choix discret aléatoire du Logit

La modélisation de choix discret aléatoire est largement utilisé dans l'étude des réseaux de transports ou aériens [3] où les utilisateurs ont le choix de plusieurs routes avec des utilités suivant la route choisie. Pour notre modèle, le système propose  $K$  classes de prix (segments) et son objectif est d'optimiser son revenu moyen, sachant que les choix des utilisateurs dépendent de leur utilité. Cette dernière est représentée par une fonction d'utilité linéaire en fonction de l'allocation pour cette classe de prix. Nous supposons que le temps d'utilisation de la ressource  $D$  est exogène au système et indépendante de l'allocation. Le prix par unité de temps par unité d'allocation pour une classe  $k$  est noté  $u_k$ , et  $c_k$  est l'allocation de la ressource pour cette classe.

La probabilité que les utilisateurs choisissent la classe  $k$  dépend du vecteur  $\underline{c}$  des allocations pour chaque classe et est notée  $\mathbb{P}_k(\underline{c})$ . Comme la fonction d'utilité est croissante en les variables de décision (on peut parler de fonction de des-utilité), les utilisateurs vont choisir la classe de prix leur offrant une utilité minimale. La probabilité que la classe  $k$  soit choisie par un individu est :

$$\mathbb{P}_k(\underline{c}) = \mathbb{P} \left( U_k(c_k) = \min_{i=1, \dots, K} U_i(c_i) \right).$$

Le problème d'optimisation du revenu moyen du fournisseur peut donc s'écrire :

$$\begin{aligned} \max_{\underline{c}} \sum_{k=1}^K D u_k c_k \mathbb{P}_k(c), & \quad (7.1) \\ \text{tel que } \sum_{k=1}^K c_k \leq C. & \end{aligned}$$

L'utilité d'un utilisateur choisissant la classe  $k$  est définie par

$$\begin{aligned} U_k(c_k) &= D u_k c_k + v_1 D + \chi_k, \\ &= V_k(c_k) + \chi_k, \end{aligned}$$

où le paramètre  $v_1$  représente le coût par unité de temps et  $\chi_k$  est un bruit aléatoire. L'utilité est alors modélisée par une variable aléatoire reflétant les incertitudes liées aux caractéristiques des alternatives. Ce problème d'optimisation dépend de la façon dont les utilisateurs choisissent leur classe de prix. L'hypothèse est généralement de considérer que les bruits  $\chi_k$  sont normalement distribuées. Nous obtenons alors le modèle de choix discret du Probit mais cette fonction de probabilité n'a pas de forme explicite. Nous considérons que les bruits  $\chi_k$  suivent des lois de Gumbel, une approximation de la loi Normale. Nous obtenons alors le modèle probabiliste de choix discret du Logit qui est largement utilisé pour modéliser un comportement aléatoire [2].

Nous supposons que les variables aléatoires  $\chi_k$  sont indépendantes et identiquement distribuées suivant une loi de Gumbel. La densité de Gumbel est définie par :

$$f(x) = \mu e^{-\mu(x-\eta)} e^{-e^{-\mu(x-\eta)}},$$

avec  $\eta \in \mathbb{R}$  un paramètre de localisation et  $\mu \in \mathbb{R}^+$ , un paramètre d'échelle. La moyenne de la distribution de Gumbel est :

$$\eta + \frac{\gamma}{\mu},$$

où  $\gamma = \lim_{n \rightarrow +\infty} \sum_{i=1}^n \frac{1}{i} - \ln(n) \approx 0.5772$  est la constante d'Euler. La variance est :

$$\sigma^2 = \frac{\pi^2}{6\mu^2}.$$

La distribution de Gumbel est une approximation de la loi Normale. Si  $\chi_1$  (resp.  $\chi_2$ ) est une variable de Gumbel de paramètres  $\eta_1$  (resp.  $\eta_2$ ) et  $\mu$ , alors la variable aléatoire  $\epsilon = (\chi_2 - \chi_1)$  suit une loi de distribution «logistique» de paramètre  $(\eta_2 - \eta_1)$  et  $\mu$ . La densité de probabilité  $g$  de la distribution «logistique» (avec  $\eta_1 = \eta_2$ ) est :

$$g(x) = \frac{\mu e^{\mu x}}{(1 + e^{-\mu x})^2},$$

avec  $\mu \in \mathbb{R}$ ,  $\mu > 0$ . On considère que le nombre de classe de prix est limité à  $K = 2$ , c'est-à-dire que le choix est alors binaire. La probabilité qu'un utilisateur choisisse la classe 1 est :

$$\begin{aligned} \mathbb{P}_1(c_1, c_2) &= \mathbb{P}[U_1 \leq U_2] \\ &= \mathbb{P}[V_1 + \chi_1 \leq V_2 + \chi_2] \\ &= \mathbb{P}[V_1 - V_2 \leq \chi_2 - \chi_1]. \end{aligned}$$

Par conséquent :

$$\begin{aligned} \mathbb{P}_1(c_1, c_2) &= \int_{V_1 - V_2}^{+\infty} g(x) dx \\ &= \left[ \frac{1}{e^{-\mu x} + 1} \right]_{V_1 - V_2}^{+\infty} \\ &= 1 - \frac{1}{e^{-\mu(V_1 - V_2)} + 1} \\ &= \frac{e^{-\mu V_1}}{e^{-\mu V_1} + e^{-\mu V_2}}. \end{aligned}$$

Dans le cas d'un choix avec plus de deux alternatives, on parle de modèle de Logit Multinominal. La probabilité qu'un individu choisisse la classe  $k$  parmi  $K$  est donnée alors par :

$$\mathbb{P}_k(\underline{c}) = \frac{e^{-\theta V_k(c_k)}}{\sum_{i=1}^K e^{-\theta V_i(c_i)}}, \quad (7.2)$$

où  $\theta$  est un réel strictement positif qui détermine le degré d'aléas dans le choix de l'utilisateur. Ce paramètre est généralement estimé par des jeux de tests sur les choix des utilisateurs.

Suivant la valeur de ce paramètre, nous pouvons isoler trois différentes fonctions de probabilité, illustrées sur la figure 7.1.

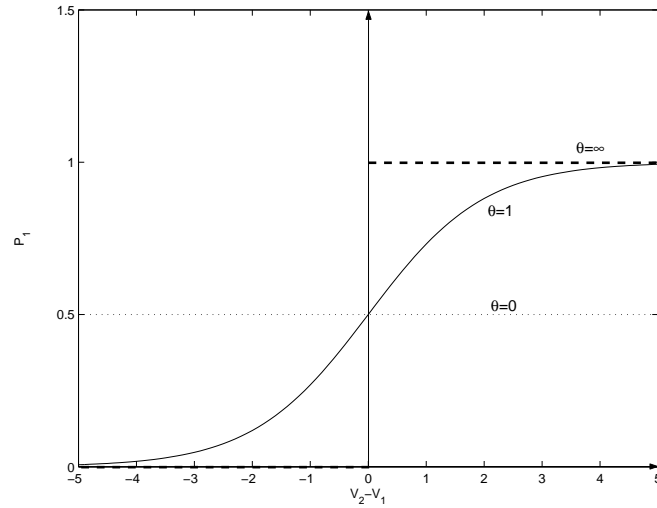


FIG. 7.1 – Probabilité de choisir la classe 1 suivant l'écart d'utilité  $V_2 - V_1$  avec le modèle du Logit avec deux classes.

1.  $\theta = 0$ . Dans ce cas, les choix sont purement aléatoires et  $\mathbb{P}_1 = \frac{1}{2}$  indépendamment des fonctions de valuations  $V_1$  et  $V_2$ .
2.  $\theta = \infty$ . Dans ce cas, le choix est purement déterministe. On obtient  $\mathbb{P}_1 = \mathbb{1}_{V_1 \leq V_2}$  qui est une fonction en escalier, non-continue en 0.
3. Quand  $\theta \in ]0, \infty[$ , la fonction de probabilité du Logit est dérivable sur son ensemble de définition. Dans ce cas de figure, le modèle prend en compte des erreurs de perception de coût aléatoires pour chacun des choix.

On peut observer lorsque  $\theta > 0$ , que la probabilité de choisir la classe 1 augmente avec l'écart d'utilité de la classe 2 par rapport à la classe 1.

### 7.3 Optimisation du revenu moyen du fournisseur

Nous analysons le modèle pour le cas particulier avec deux classes de prix, i.e.  $K = 2$ . Les prix pour chaque classe  $u_1$  et  $u_2$  sont fixés par le système suivant des études de marché par exemple. Le système optimise les allocations  $c_1$  et  $c_2$  pour chaque classe de prix ou segment de marché afin de maximiser son revenu moyen. On suppose que le choix de la classe des prix par les utilisateurs se modélise par un modèle de choix discret du Logit.

D'après l'expression des probabilités (7.2), le problème d'optimisation du revenu moyen (7.1) s'écrit :

$$\max_{c_1, c_2 \geq 0} D \left( \frac{u_1 c_1}{1 + e^{\theta D(u_1 c_1 - u_2 c_2)}} + \frac{u_2 c_2}{1 + e^{\theta D(u_2 c_2 - u_1 c_1)}} \right),$$

tel que  $c_1 + c_2 \leq C$ .

Le revenu est maximisé si la totalité de la ressource  $C$  est allouée, ainsi la contrainte d'inégalité du problème d'optimisation est active à l'optimum, c'est-à-dire que l'on doit avoir  $c_1^* + c_2^* = C$ . Dans ce cas, nous simplifions l'analyse en considérant une unique variable de décision  $c_1$ . Ainsi, la fonction objectif peut s'écrire seulement en fonction de  $c_1$  par :

$$F(c_1) = D(f(c_1) + g(c_1)),$$

avec

$$f(c_1) = \frac{u_1 c_1}{1 + e^{\theta D(u_1 c_1 - u_2(C - c_1))}}$$

et

$$g(c_1) = \frac{u_2(C - c_1)}{1 + e^{\theta D(u_2(C - c_1) - u_1 c_1)}}.$$

Ce problème d'optimisation avec seulement deux classes de prix peut être résolu de façon analytique. Pour cela nous définissons les fonctions suivantes

$$y(c_1) = e^{\theta D c_1(u_1 + u_2) - \theta D u_2 C},$$

et

$$M(c_1) = \frac{1}{\sqrt{u_2}} \sqrt{\frac{H(c_1)^2}{4u_2} + u_1} - \frac{H(c_1)}{2u_2},$$

avec  $H(c_1) = u_2 - u_1 + \theta D(u_1 + u_2)(u_1 c_1 - u_2(C - c_1))$ .

**Théorème 20** Soit  $(c_1^*, c_2^*)$  l'allocation optimale de la ressource avec deux classes de prix, i.e.  $K = 2$ .

– Si  $y(C) \leq M(C)$ , c'est-à-dire

$$e^{\theta D C u_1} \leq \frac{1}{\sqrt{u_2}} \sqrt{\frac{(u_2 - u_1 + \theta D(u_1 + u_2)(u_1 C))^2}{4u_2} + u_1} - \frac{(u_2 - u_1 + \theta D(u_1 + u_2)(u_1 C))}{2u_2},$$

alors l'allocation optimale est  $(c_1^*, c_2^*) = (C, 0)$ .

– si  $y(0) \geq M(0)$ , c'est-à-dire

$$e^{-\theta D u_2 C} \geq \frac{1}{\sqrt{u_2}} \sqrt{\frac{(u_2 - u_1 - \theta D(u_1 + u_2)(u_2 C))^2}{4u_2} + u_1} - \frac{(u_2 - u_1 - \theta D(u_1 + u_2)(u_2 C))}{2u_2},$$

alors l'allocation optimal est  $(c_1^*, c_2^*) = (0, C)$ .

– Si  $y(C) \geq M(C)$  et  $y(0) \leq M(0)$ , l'allocation optimale est  $(c_1^*, c_2^*) = (x^*, C - x^*)$ , avec  $x^*$  l'unique solution de  $y(x) = M(x)$ .

**Preuve** D'après l'expression de la fonction  $y(c_1)$ , on peut réécrire la fonction objectif  $F(c_1)$  sous la forme :

$$F(c_1) = D \left( \frac{u_1 c_1}{1 + y(c_1)} + \frac{u_2(C - c_1)}{1 + 1/y(c_1)} \right).$$

Sa dérivée par rapport à  $c_1$  est

$$F'(c_1) = D \left( \frac{u_1}{(1+y)^2} (1+y - c_1 y') - y^2 \frac{u_2}{(1+y)^2} (1+y^{-1} - (C - c_1) y' / y^2) \right).$$

La fonction  $F$  est croissante si et seulement si

$$y^2 u_2 + y(u_2 - u_1) + y'(u_1 c_1 - u_2(C - c_1)) \leq u_1. \quad (7.3)$$

Or, on a

$$y'(c_1) = \theta \zeta_1 D(u_1 + u_2) y(c_1),$$

et donc l'équation (7.3) devient l'équation du second degré fonctionnelle suivante

$$y^2 u_2 + y H(c_1) \leq u_1,$$

avec  $H(c_1) = u_2 - u_1 + \theta D(u_1 + u_2)(u_1 c_1 - u_2(C - c_1))$ . On obtient donc que la fonction objectif est croissante si et seulement si

$$y\sqrt{u_2} \leq \sqrt{\frac{H(c_1)^2}{4u_2} + u_1} - \frac{H(c_1)}{2\sqrt{u_2}}.$$

Cette dernière équation est équivalente à :

$$y(c_1) = e^{\theta D c_1 (u_1 + u_2) - \theta D u_2 C} \leq \frac{1}{\sqrt{u_2}} \sqrt{\frac{H(c_1)^2}{4u_2} + u_1} - \frac{H(c_1)}{2u_2} = M(c_1).$$

La fonction  $y(\cdot)$  est continue et strictement croissante et la fonction  $M(c_1)$  est strictement décroissante. En effet, sa dérivée est égale à :

$$\left( \frac{1}{\sqrt{u_2}} \sqrt{\frac{H(c_1)^2}{4u_2} + u_1} - \frac{H(c_1)}{2u_2} \right)' = \frac{\theta D (u_1 + u_2)^2}{2u_2} \left( \frac{H(c_1)}{\sqrt{H(c_1)^2 + 4u_1 u_2}} - 1 \right) < 0.$$

Il suffit alors de comparer la valeur des fonctions aux bords du domaine d'optimisation  $[0, C]$  afin de déterminer l'allocation optimale. Nous avons 3 cas de figures possibles.

1. Si  $y(0) \geq M(0)$ , nous avons, pour tout  $c_1 \in [0, C]$ ,  $y(c_1) \geq M(c_1)$  et donc la fonction objectif est strictement décroissante en  $c_1$ , et l'allocation optimale est donc  $(c_1^*, c_2^*) = (0, C)$ .
2. Si  $y(C) \leq M(C)$  alors la fonction objectif est strictement croissante en  $c_1$  sur l'intervalle des allocations  $[0, C]$ , et donc l'allocation optimale est  $(c_1^*, c_2^*) = (C, 0)$ .
3. Si  $y(0) \leq M(0)$  et  $y(C) \geq M(C)$ , alors les deux courbes se croisent sur l'intervalle  $[0, C]$ , il existe donc un unique réel  $x^* \in ]0, C[$  tel que  $y(x^*) = M(x^*)$ , et l'allocation optimale est donc  $(c_1^*, c_2^*) = (x^*, C - x^*)$ .

■

Nous illustrons ce théorème sur un exemple puis appliquons notre modèle à des données réelles correspondants à des requêtes sur des serveurs web.

*Exemple* La figure 7.2 représente un exemple d'allocation optimale pour  $K = 2$  classes de prix et avec comme paramètres : le temps d'utilisation de la ressource  $D = 2$ , le paramètre d'aléas de la fonction Logit  $\theta = 0.05$ , les prix pour chaque classe  $u_1 = 3$ ,  $u_2 = 6$  et la capacité totale à allouer  $C = 10$ . Nous vérifions que  $y(10) > M(10)$  et  $y(0) < M(0)$  donc l'allocation optimale donnée par le théorème 20 est  $c_1^* = 6.2892$  et  $c_2^* = 3.7108$ .

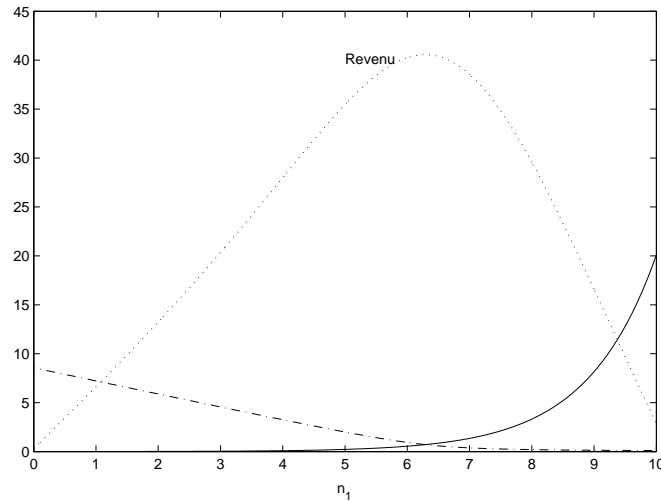


FIG. 7.2 – Allocation optimale de la ressource pour la première classe avec le modèle de choix discret du Logit et 2 classes de prix.

L'extension avec 3 classes de prix est beaucoup plus complexe car nous avons un système d'équations non-linéaires à résoudre. La figure 7.3 montre un exemple d'optimisation avec 3 classes de prix. Nous considérons les paramètres suivants :  $\theta = 0.05$ ,  $\zeta_1 = 1$ ,  $u_1 = 2$ ,  $u_2 = 4$ ,  $u_3 = 6$ ,  $C = 10$  et  $D = 4$ . La solution est  $c_1^* = 4.9678$ ,  $c_2^* = 2.9294$  et  $c_3^* = 2.1027$ . Le revenu moyen du fournisseur est ni concave, ni quasi-concave car les courbes de niveau ne sont pas convexes (voir figure 7.4). Cependant, c'est une fonction non-convexe dont l'enveloppe lisse permet à n'importe quel algorithme standard d'optimisation basé sur la méthode du gradient de converger vers l'optimum.

Dans la suite, nous appliquons notre modèle sur des données réelles collectées sur un serveur web. Ces données représentent la demande à chaque instant, ou plus exactement la quantité d'informations demandée par les utilisateurs. Nous étudions le revenu moyen du fournisseur lorsque celui-ci propose plusieurs classes de prix et lorsque les utilisateurs choisissent leur classe suivant un modèle du Logit. Nous comparons le revenu moyen suivant le nombre de classes (segments) de prix.

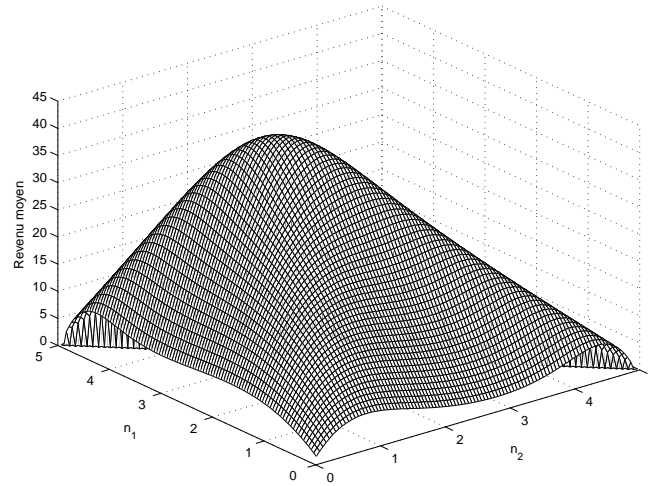


FIG. 7.3 – Revenu moyen du fournisseur avec 3 classes et modèle de choix discret du Logit.

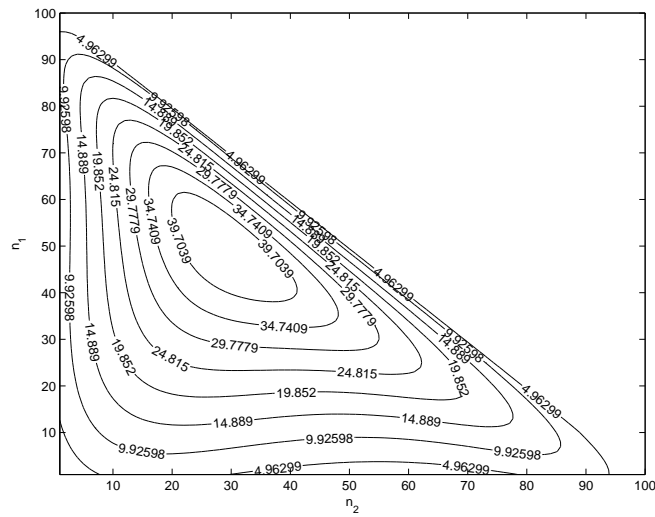


FIG. 7.4 – Courbes de niveau du revenu moyen avec 3 classes de prix et modèle de choix discret du Logit.



## 7.4 Application à des requêtes sur un serveur web

Nous illustrons notre analyse précédente sur des données numériques de requêtes sur un serveur web. Nous observons la quantité de données transmises sur 8 jours. Les données ne prennent pas en compte la durée de travail et nous considérons donc une durée constante  $D = 1$  arbitraire. La demande représente la ressource totale à répartir entre les différentes classes de prix. L'objectif est alors de déterminer l'allocation ou répartition optimale pour chaque classe afin d'optimiser le revenu moyen. Nous donnons le revenu optimal (avec allocation optimale de la ressource) en fonction du nombre de classe de prix. Nous le comparons avec le cas de base comprenant un prix unique et montrons que le revenu moyen obtenu est alors meilleur.

Les données transmises représentent la demande pour chaque instant. Le modèle permet qu'un utilisateur choisisse de ne pas rentrer dans le système en utilisant un prix très petit (nous avons utilisé  $u_0 = 0.01$ ). On limite ainsi les prix car cette alternative est choisie avec une forte probabilité si les autres prix sont très grands. Les autres prix sont déterminés arbitrairement dans le tableau 7.4 où l'on considère un nombre variable de prix, de 1 à 6. Dans le cas d'un unique prix, on considère 3 cas de figure : un prix faible, moyen ou élevé. Pour le cas avec plusieurs classes de prix, nous les répartissons entre 0 et 1 plus ou moins uniformément. Ces prix servent juste à justifier le modèle et dans la réalité, ils peuvent être fixés sur la base d'études au préalable de marché.

$K$	Prix
faible	.2
moyen	.6
élevé	1
2	.4 .8
3	.3 .6 .9
4	.2 .4 .6 .8
5	.2 .4 .6 .8 1
6	.2 .35 .5 .65 .8 .95

TAB. 7.1 – Tableau des prix pour chaque classe.

La figure 7.5 montre le revenu optimal au cours du temps en considérant de 2 à 6 classes de prix. La courbe supérieure représente la demande totale. On observe qu'un grand nombre de classes de prix (5-6) donne clairement un revenu optimal plus important qu'un petit nombre de segments en période de demande importante. Alors que pendant les périodes de faible demande, seulement 2 ou 3 classes de prix donne un meilleur revenu moyen.

La figure 7.6 illustre le revenu maximum à des instants précis où la demande est différente : deux instants avec demande faible (les instants 40 et 200), un instant avec demande moyenne (instant 80) et deux instants avec demande forte (les instants 120 et 160). Ceci nous montre que le nombre de segments qui optimise le revenu moyen

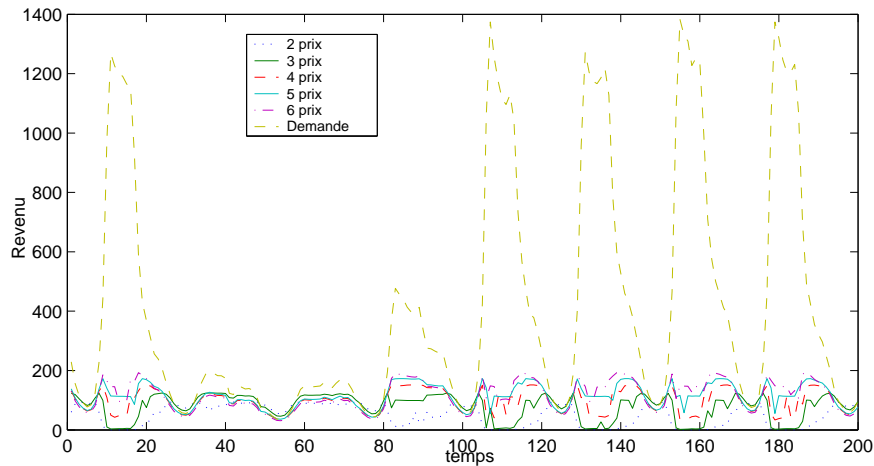


FIG. 7.5 – Revenu optimal avec plusieurs classes de prix.

varie suivant la demande totale. Et précisément, plus la demande est importante, plus le nombre de classes de prix doit être important.

En effet, on remarque pour les périodes avec une demande importante, comme par exemple les instants 120 et 160, le revenu est optimum avec un grand nombre de classes de prix (5 ou 6). Inversement, quand la demande est faible, par exemple aux instants 40 et 200, le revenu est optimum avec seulement 2 ou 3 classes de prix. Ces observations peuvent s'expliquer à cause de la séparation de la demande entre les classes de service. Si la demande est faible, la probabilité d'avoir de choisir les classes les plus chers est faible.

Nous comparons maintenant le revenu optimal avec un unique prix (faible, moyen ou élevé) et une stratégie qui consiste à proposer 5 classes de prix différents.

On remarque tout d'abord la stratégie avec 5 classes de prix donne un revenu optimal toujours supérieur au cas d'un prix unique, indépendamment du fait que celui-ci soit faible, moyen ou élevé (à part les instants 55 et 78 où la demande est très faible mais ces instants sont rares).

Enfin sur la figure 7.8 nous représentons l'allocation optimale de chaque classe de prix lorsque la stratégie du fournisseur est de proposer 5 prix différents. Nous remarquons que cette allocation varie en fonction de la demande. Plus elle est importante, plus l'allocation des prix élevé est importante. Ceci implique que le mécanisme d'allocation optimale doit être relancer à chaque instant et pose donc le problème de granularité.

## 7.5 Conclusion

Ce chapitre a porté sur l'étude d'un mécanisme optimal de réservation de ressources dans le but d'optimiser le revenu moyen du fournisseur. Le comportement aléatoire des utilisateurs est basé sur le modèle de choix discret du Logit. Ce dernier est souvent

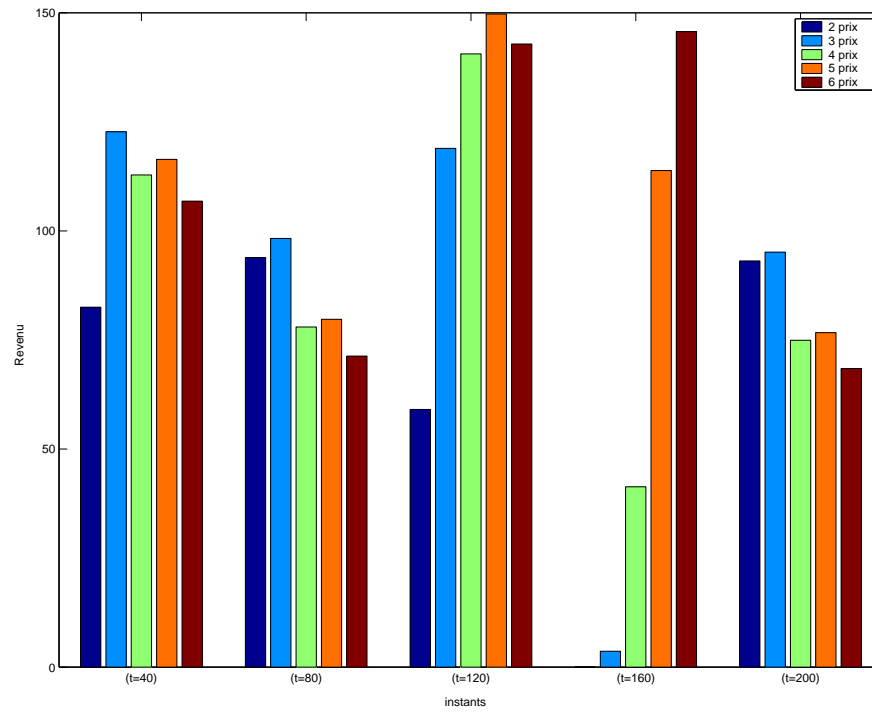


FIG. 7.6 – Revenu optimal pour 5 instants différents (périodes avec demande faible(40), moyenne (80), forte (120), forte (160) et faible (200)) avec les 5 stratégies proposées (de 2-6 prix).

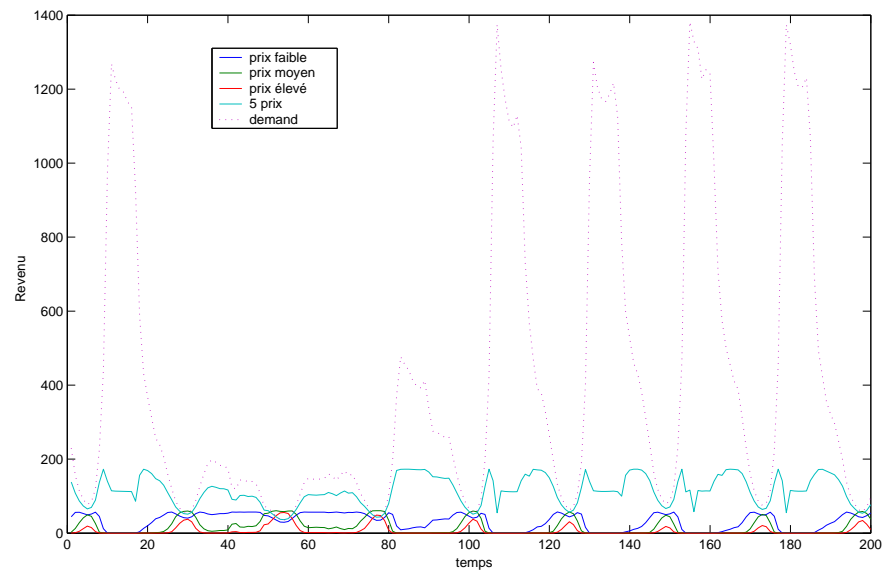


FIG. 7.7 – Comparaison du revenu optimal avec 5 prix et un prix unique, faible, moyen et élevé.

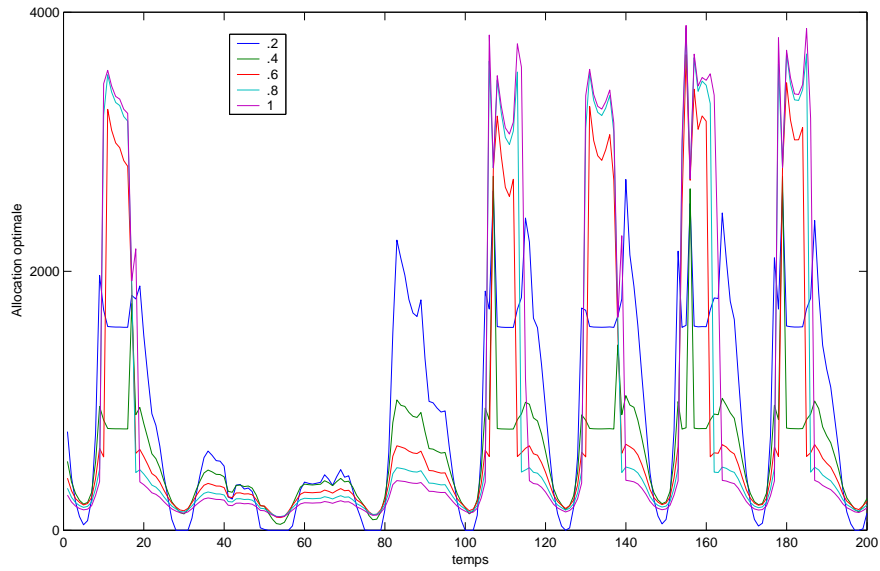


FIG. 7.8 – Allocation optimale pour chacune des 5 classes de prix.

utilisé pour la modélisation de choix d'utilisateurs dans les réseaux transports et n'a jamais été utilisé à notre connaissance avec un mécanisme de *yield management*. Nous avons tout d'abord proposé une étude analytique du problème d'optimisation avec un faible nombre de classes de prix et des résultats numériques lorsque ce nombre devient plus important. Enfin, nous avons appliqué le modèle à des données réelles de transaction web.

Cette dernière étude a permis de mettre en œuvre une technique de *yield management* pour l'optimisation d'un nouveau modèle de tarification de l'Internet.

## Chapitre 8

# Conclusion générale

L'avènement de nouvelles applications de type multimédia (VoIP, VoD, ...) très demandeuses en qualité de service, et le nombre croissant permanent d'utilisateurs font que l'Internet nouvelle génération sera fortement propice au problème de congestion. Cette congestion sera particulièrement ressentie au niveau des réseaux limités en capacité comme les réseaux d'accès et les réseaux sans fil. Plusieurs architectures de réseaux basées sur une notion de différenciation de service ont été proposées par l'IETF afin de répondre aux critères de qualité de service des nouvelles applications, mais ces architectures ne contrôlent pas la demande qui est le générateur principal de congestion. Un mécanisme simple de contrôle de la demande est d'utiliser une tarification à l'usage qui incite les utilisateurs à contrôler leur consommation et donc leur trafic. Nous nous sommes donc intéressés dans cette thèse à l'étude (modélisation et optimisation) de plusieurs mécanismes de tarification à l'usage pour différents modèles de réseaux à différenciation de services.

### 8.1 Contributions

Nous avons considéré plusieurs modèles de réseaux, et pour chacun d'eux, nous avons étudié le comportement des utilisateurs face à une tarification à l'usage. Nous avons également optimisé les paramètres du mécanisme de tarification, en considérant le comportement des utilisateurs induit par les prix et ainsi nous proposons au fournisseur une tarification optimale des services offerts sur son réseau. Le fournisseur décide de ses prix suivant deux objectifs : l'optimisation du surplus collectif ou l'optimisation de son revenu.

Nous nous sommes intéressés tout d'abord au modèle du *Cumulus*. Ce mécanisme de tarification basé sur un contrat déterminant la quantité de ressources entre chaque utilisateur et son fournisseur d'accès. Le *Cumulus* est prometteur de par sa simplicité de mise en œuvre sur une architecture de réseau à différenciation de services de type *DiffServ*. Notre première contribution a été de proposer une modélisation du comportement des utilisateurs et du revenu du fournisseur, face à ce nouveau modèle de tarification. Ensuite, nous avons déterminé comment le fournisseur doit positionner cer-

tains paramètres fondamentaux du modèle pour lesquels le choix était arbitraire dans la littérature.

Le modèle du Cumulus est fortement lié à l'architecture de réseaux *DiffServ* dans laquelle la différenciation de service se fait au niveau des routeurs, par un traitement différencié des paquets suivant leur classe. Plusieurs algorithmes d'ordonnancement ont été standardisés par la communauté scientifique. Dans le chapitre suivant, nous nous sommes donc intéressés à comparer, d'un point de vue économique, différents mécanismes d'ordonnancement. Nous avons particulièrement montré que le mécanisme d'ordonnancement avec priorités strictes engendre un meilleur revenu pour le fournisseur de service qu'une approximation du mécanisme à processeur partagé généralisé (GPS) et que le mécanisme à processeur partagé discriminant (DPS). L'algorithme GPS porte un grand intérêt dans la communauté scientifique car il assure un délai minimal à chaque classe de service. Quant à l'algorithme DPS, il permet de modéliser le partage de ressource de multiples connections TCP en concurrence à un nœud du réseau. Ainsi, proposer de la différenciation de service dans les réseaux permet d'une part d'engendrer des bénéfices pour le fournisseur et d'autre part, de proposer une meilleure qualité de service aux utilisateurs que le *best effort* actuel. La conclusion importante est que le mécanisme à priorités strictes donne un meilleur revenu au fournisseur qu'un algorithme de partage de ressource équitable comme GPS pour un modèle de réseau M/M/1 et pour un modèle de connections TCP hétérogènes en concurrence (modélisé avec DPS).

De nombreux mécanismes de tarification de la littérature portent sur des systèmes à différenciation de service avec un mécanisme d'ordonnancement à priorités strictes. Les principaux modèles, basés sur le modèle de Mendelson et Whang [81], proposent une tarification au coût marginal optimale au sens du surplus collectif, mais qui ne partage pas le coût total de congestion du système. Dans le chapitre 4, nous proposons alors un mécanisme qui, en plus d'être optimal au sens du surplus collectif, possède la propriété de partager le coût de congestion du système entre les classes de service. Nous avons ensuite proposé une extension de ce mécanisme à un modèle de tarification dynamique, où la demande varie dans le temps, et enfin à un modèle de réseau ouvert de Jackson.

Un fournisseur proposant un mécanisme de tarification à l'usage doit nécessairement mesurer la consommation des utilisateurs de son réseau (supposée connue jusqu'alors dans les modèles de la littérature). Dans le chapitre 5, nous supposons que le système estime par le biais de mesures actives et passives les performances du réseau pour chaque classe de service. Le système utilise des techniques de mesure active, en injectant un trafic afin d'estimer le délai moyen et de mesure passive, en observant une proportion du trafic traité, afin d'estimer le débit moyen de chaque classe de service. La mise en œuvre de ces mécanismes influe sur le comportement des utilisateurs qui sont supposés sensibles aux erreurs de mesure effectuées par le système. Nous avons déterminé, pour ce nouveau modèle, les prix et les paramètres de mesures qui maximisent le surplus collectif du système lequel prend en compte les coûts des mesures. Nous avons ensuite proposé plusieurs algorithmes d'estimation des mesures de performances qui diffèrent par le poids d'estimation des mesures dans le passé pour le nouvel estimateur.

La qualité de service et le contrôle de congestion sont également des problèmes importants dans les réseaux sans fil. Et plus encore, dans les réseaux sans fils de troisième

génération UMTS basés sur la technologie CDMA qui proposent des services temps-réels comme la visiophonie, la télévision, etc. Ce type de réseau est très sujet au problème de congestion à cause de la limitation du débit par le spectre radio. Nous étudions plus particulièrement les transmissions sur la voix montante avec plusieurs types d'applications qui diffèrent par leur sensibilité à leur qualité de service. Nous avons proposé une tarification statique optimale et nous avons montré que, lorsque la demande excède toujours la capacité, le fournisseur optimise son revenu avec une unique classe d'application. Alors que lorsque la demande est aléatoire, le revenu moyen du fournisseur est optimisé avec plusieurs classes de service.

L'optimisation du revenu pour des systèmes avec demande aléatoire possède un grand intérêt dans la communauté des réseaux et est connue sous le nom de *yield management*. Ce mécanisme marketing permet de calculer les meilleurs prix pour optimiser le profit généré par la vente d'un produit ou d'un service, sur la base d'une modélisation et d'une prévision du comportement de la demande par micro segment du marché. Dans le dernier chapitre, nous appliquons un modèle de *yield management* afin de déterminer une allocation optimale de ressources entre plusieurs classes de service où les prix sont fixés. Le *yield management* est une technique qui vient directement du marketing et qui est très peu utilisé dans un contexte de tarification des réseaux de communication.

## 8.2 Perspectives

Les perspectives de mes travaux de thèse sont nombreuses compte tenu de la diversité des modèles étudiés dans cette thèse.

- Nous pensons généraliser notre modèle de tarification optimale avec partage de coût pour d'autres formes de fonction d'utilité comme celle de Cobb-Douglas.
- Nous travaillons à l'extension de notre modèle de tarification combiné avec des techniques de mesures actives et passives, avec une politique d'ordonnancement à priorités strictes entre classes de service.
- Nous envisageons d'étudier l'impact que peut avoir la méthodologie de mesure de consommation des utilisateurs pour le modèle du Cumulus. Un problème intéressant de ce modèle porte sur les différents facteurs d'échelle de temps, la granularité.
- Il serait également intéressant de comparer expérimentalement les algorithmes d'ordonnancement à priorités strictes avec des émulations à paquets de l'algorithme GPS comme WFQ ou WRR. On envisage également de prendre en compte d'autres paramètres de qualité de service comme la probabilité de perte pour un modèle à buffer fini. Nous pensons également étudier et comparer les politiques d'ordonnancement avec différents types d'applications et particulièrement des applications multimédia.
- Notre étude portant sur les réseaux sans fils CDMA porte un grand intérêt pour nos travaux futurs. Une extension à un modèle d'envoi aléatoire des paquets est en cours, nous envisageons également l'extension de notre modèle à la tarification

du trafic sur la voix descendante où la station de base adapte les puissances d'émission à chaque utilisateurs comme dans [70].

- Enfin, l'utilisation du *yield management* pour la tarification des réseaux de communication est une méthodologie récente et très peu étudiée. Il existe donc de nombreuses perspectives pour notre travail élaboré dans le chapitre 7. Nous envisageons par exemple de prendre en considération un modèle de file d'attente et dont le temps de séjour des paquets dépend du choix des utilisateurs et inversement.

Les modèles étudiés dans cette thèse traitent d'une tarification statique car, la facture étant prévisible, ce type de tarification est préféré par les utilisateurs [105]. Néanmoins, une tarification dynamique dont les prix évoluent en fonction de la demande, est plus efficace et est un bon moyen pour le fournisseur d'optimiser ses bénéfices. Son efficacité repose sur la souplesse du système de tarification. Nous envisageons d'étudier des modèles de tarification dynamique pour un réseau à différenciation de services comme dans [40] où l'évolution du système est décrite par un processus de décision markovien [91]. Il serait également intéressant d'étudier des mécanismes de tarification pour les réseaux ad-hoc afin d'inciter les utilisateurs à transmettre les paquets [33].

En dehors d'un contexte de tarification, la gestion de la qualité de service dans les réseaux à capacité limitée et particulièrement les réseaux sans fil (cellulaires ou ad-hoc) porte un grand intérêt pour mes futurs travaux de recherche. J'envisage par exemple d'étudier les mécanismes d'ordonnancement mis en place dans une cellule CDMA comme dans [65], ou encore des mécanismes de différenciation à la *DiffServ* pour les réseaux ad-hoc proposés par la norme 802.11e [115].



# Bibliographie

- [1] Long-term traffic statistics. <http://www.cs.columbia.edu/~hgs/internet/traffic.html>.
- [2] B. Akiva and S. Lerman. Discrete Choice Analysis : Theory and Application to Travel Demand. Technical report, MIT Press, 1985.
- [3] M. Akiva and M. Bierlaire. *Handbook of Transportation Science*, chapter Discrete Choice Methods and Their Application to Short Travel Decisions. Kluwer, 1999.
- [4] M. Allman, V. Paxson, and W. Stevens. TCP Congestion Control. IETF RFC 2581, 1999.
- [5] T. Alpcan, T. Basar, R. Srikant, and E. Altman. CDMA Uplink Power Control as a Noncooperative Game. *Wireless Networks*, 8, 2002.
- [6] E. Altman, C. Barakat, and V. Ramos. Queuing Analysis of Simple FEC Schemes over IP. *Computer Networks*, 39(2), 2002.
- [7] E. Altman, T. Jimenez, and Kofman D. DPS Queues with Stationary Ergodic Service Times and the Performance of TCP in Overload. In *Proceedings of INFOCOM*, 2004.
- [8] E. Altman and L. Wynter. Equilibrium, Games, and Pricing in Transportation and Telecommunication Networks. *Networks and Spatial Economics*, 4, 2004.
- [9] L. Anania and R. Solomon. *Internet Economics*, chapter Flat- The Minimalist Price, pages 91–118. MIT Press, 1997.
- [10] S. Ariyavisitakul and L. Chang. Signal and Interference Statistics of a CDMA System with Feedback Power Control. *IEEE Transactions on Communications*, 41(11), 1993.
- [11] K. Avrachenkov, U. Ayesta, P. Brown, and R. Núñez Quieja. Discriminatory Processor Sharing Revisited. In *Proceedings of IEEE Infocom*, 2005.
- [12] T. Basar and G. Olsder. *Dynamic Noncooperative Game Theory*. SIAM Series in Classics in Applied Mathematics, 1999.
- [13] T. Basar and R. Srikant. Revenue-Maximizing Pricing and Capacity Expansion in a Many-Users Regime. In *Proceedings of INFOCOM*, 2002.

- 
- [14] W. Beckert. *Stochastic Demand Analysis*. PhD thesis, UC Berkeley, 2000.
- [15] P. Belobaba. Airline Yield Management : an Overview of Seat Inventory Control. *Transportation Science*, 21, 1987.
- [16] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, and J. Roberts. Statistical Bandwidth Sharing : a Study of Congestion at Flow Level. In *Proceedings of IEEE SIGCOMM*, 2001.
- [17] L. Bernstein. Managing the Last Mile. *IEEE Communications Magazine*, 35(10) :72–76, 1997.
- [18] L. Billera, D. Heath, and J. Raanan. Internal Telephone Billing Rates - A Novel Application of Non-Atomic Game Theory. *Operations Research*, 26(6) :956–965, 1978.
- [19] J. Bolot. Characterizing End-to-End Packet Delay and Loss in the Internet. *Journal of High Speed Networks*, 2, 1993.
- [20] J. Bolot. End-to-End Packet Delay and Loss Behavior in the Internet. In *Proceedings of ACM SIGCOMM*, 1993.
- [21] J. Bonnin, B. Jechoux, P. Jolivet, X. Lagrange, P. Martins, L. Nuaymi, and S. Tabbane. *Principes et Evolutions de l'UMTS*. Hermes, 2005.
- [22] S. Borst, O. Boxma, and R. Núñez Queija. The Effects of the Service Discipline. In *Proceedings of Performance TOOLS*, 2002.
- [23] R. Braden, D. Clark, and S. Shenker. Technical report.
- [24] R. Bradford. Pricing, Routing and Incentive Compatibility in Multiserver Queues. *European Journal of Operational Research*, 89, 1996.
- [25] L. Brotcorne, M. Labbe, P. Marcotte, and G. Savard. A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network. *Transportation Science*, 35, 2001.
- [26] T. Bu and D. Towsley. Fixed Point Approximation for TCP Behavior in an AQM Network. In *Proceedings of ACM SIGMETRICS*, 2001.
- [27] X. Cao and H. Shen. Internet Pricing : Comparison and Examples. In *Proc. 39th IEEE Conf. Decision and Control*, pages 2284–2289, 2000.
- [28] D. Clark and W. Fang. Explicit Allocation of Best Effort Packet Delivery Service. *IEEE/ACM Transactions on Networking*, 6(4), 1998.
- [29] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang. Pricing in Computer Networks : Motivation, Formulation, and Example. *IEEE/ACM Transactions on Networking*, 1(6) :614–627, 1993.

- 
- [30] J. Cohen. The Multiple Phase Service Network with Generalized Processor Sharing. *Acta Informatica*, 12, 1979.
- [31] C. Courcoubetis, V. Siris, and G. Stamoulis. Integration of Pricing and Flow Control for Available Bit Rate Services in ATM Networks. In *Proceedings of IEEE Globecom*, 1996.
- [32] C. Courcoubetis and R. Weber. *Pricing Communications Networks*. Wiley, 2003.
- [33] J. Crowcroft, R. Gibbens, F. Kelly, and S. Ostring. Modelling Incentives for Collaboration in Mobile Ad-Hoc Networks. *Performance Evaluation*, 57, 2004.
- [34] L.A. DaSilva. Pricing of QoS-Enabled Networks : A Survey. *IEEE Communications Surveys & Tutorials*, 3(2), 2000.
- [35] A. Demers, S. Keshav, and S. Shenker. Analysis and Simulation of a Fair Queueing Algorithm. *Internetworking : Research and Experience*, 1 :3-26, 1990.
- [36] K. Deschinkel. *Régulation du trafic aérien par optimisation dynamique des prix d'utilisation du réseau*. PhD thesis, ONERA Centre de Toulouse, 2001.
- [37] T. Dolan. Internet Pricing : Is the end of the world wide wait in view ? *Communications & Strategies*, (37) :15-46, 2000.
- [38] A. Downey. Using Pathchar to Estimate Internet Link Characteristics. In *Proceedings of ACM SIGCOMM*, 1999.
- [39] P. Dube, Y. Hayel, and L. Wynter. Yield Management for IT Resources on demand : Analysis and Validation of a New Paradigm for Managing Computing Centers. *Journal of Revenue and Pricing Management*, 4(1), 2005.
- [40] P. Dube, Borkar V., and D. Manjunath. Differential Join Prices for Parallel Queues : Social Optimality, Dynamic Pricing Algorithms and Application to Internet Pricing. In *proceedings of IEEE INFOCOM*, 2002.
- [41] R. Edell and P. Varaiya. Providing Internet Access : What We Learn From INDEX. In *Infocom'99*, 1999.
- [42] D. Famolari, N. Mandayam, D. Goodman, and V. Shah. *Wireless Multimedia Network Technologies*, chapter A New Framework for Power Control in Wireless Data Networks : Games, Utility and Pricing. Kluwer, 1999.
- [43] G. Fayolle, I. Mitrani, and R. Iasnogorodski. Sharing a Processor Among Many Job Classes. *Journal of the Association for Computing Machinery*, 27(3), 1980.
- [44] S. Floyd and V. Jacobson. Random Eraly Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1(4), 1993.
- [45] D. Fudenberg and J. Tirole. *Game Theory*. The MIT Press, 1992.

- 
- [46] ITU-T Recommendation G.107. The E-model, a Computational Model for Use in Transmission Planning, 2003.
- [47] R. Gibbens, R. Mason, and R. Steinberg. Internet Service Classes Under Competition. *IEEE Journal on Selected Areas in Communications*, 18(12), 2000.
- [48] S. Glisic and B. Vucetic. *Spread Spectrum CDMA Systems for Wireless Communications*. Artech House, 1997.
- [49] G. Guthrie and M. Carter. User Charges for Internet : The New Zealand Experience. *Telecommunication Systems : Modeling Analysis Design and Management*, 6, 1996.
- [50] R. Hassin and M. Haviv. *To Queue or Not To Queue*. Kluwer's, 2003.
- [51] M. Haviv. The Aumann-Shapley Price Mechanism for Allocating Congestion Costs. *Operations Research Letters*, 29(5) :211–215, 2001.
- [52] Y. Hayel, D. Ros, and B. Tuffin. Less-than-Best-Effort Services : Pricing and Scheduling. In *proceedings of IEEE INFOCOM*, 2004.
- [53] Y. Hayel and B. Tuffin. A mathematical analysis of the cumulus pricing scheme. *Computer Networks*, 47(6), 2004.
- [54] Y. Hayel and B. Tuffin. Pricing for Heterogeneous Services at a Discriminatory Processor Sharing Queue. In *Proceedings of IFIP/TC6 Networking Conference*, 2005.
- [55] H. Holma and A. Toksala. *WCDMA for UMTS*. John Wiley and Sons, 2001.
- [56] E. Hong, S. Hwang, K. Kim, and K. Whang. Synchronous Transmission Technique for the Reverse Link in DS-CDMA Terrestrial Mobile Systems. *IEEE Transactions on Communications*, 47(11), 1999.
- [57] L. Ingber. Simulated Annealing : Practice versus Theory. *Mathematical Computing Modelling*, 18(11) :29–57, 1993.
- [58] F. Kelly. Charging and Rate Control for Elastic Traffic. *European Transactions on Telecommunications*, 8, 1997.
- [59] F. Kelly. Congestion Control : Fairness, Pricing and Stability. In *15th IEEE Computer Communications Workshop*, 2000.
- [60] F. Kelly, A. Maullo, and D. Tan. Rate Control for Communication Networks : Shadow Prices, Proportionnal Fairness and Stability. *Journal of Operation Research Society*, 49 :237–252, 1998.
- [61] J. Kim and M. Honig. Resource Allocation for Multiple Classes of DS-CDMA Traffic. *IEEE Transactions on Vehicular Technology*, 49(2), 2000.

- [62] L. Kleinrock. Time-shared Systems : A Theoretical Treatment. *Journal of the Association for Computing Machinery*, 14(2), 1967.
- [63] L. Kleinrock. *Queueing Systems : Volume I, Theory*. John Wiley & Sons, 1975.
- [64] L. Kleinrock. *Queueing Systems : Volume II, Computer Applications*. John Wiley & Sons, 1976.
- [65] I. Koukoutsidis, E. Altman, and J.M. Kelif. Fair Rate Sharing Models in a CDMA Link with Multiple Classes of Elastic Traffic. Technical report, INRIA, 2005.
- [66] A. Kumar, K. Hari, R. Shobhanjali, and s. Sharma. Long Range Dependence in the Aggregate Flow of TCP Controlled Elastic Sessions : An Investigation via the Processor Sharing Model. In *Proceedings of National Conference on Communications*, 2000.
- [67] S. Kunniyur and R. Srikant. End-to-End Congestion Control Schemes : Utility Functions, Random Losses and ECN Marks. *IEEE/ACM Transactions on Networking*, 11(5), 2003.
- [68] K. Lai and M. Baker. Nettimer : a Tool for Measuring Bottleneck Link Bandwidth. In *Proceedings of 3rd USENIX Symposium on Internet Technologies and Systems*, 2001.
- [69] J. Lee, R. Mazumdar, and N. Shroff. Downlink Power Allocation for Multi-class CDMA Wireless Networks. In *Proceedings of IEEE INFOCOM*, 2002.
- [70] P. Liu, M. Honig, and S. Jordan. Forward-Link CDMA Resource Allocation Based on Pricing. In *Proceedings of IEEE WCNC*, 2000.
- [71] S. Low. Optimization Flow Control with On-line Measurement or Multiple Paths. In *Proceedings of the 16th International Teletraffic Congress*, 1999.
- [72] S. Low and D. Lapsley. Optimization Flow Control, I : Basic Algorithm and Convergence. *IEEE/ACM Transactions on Networking*, 7(6), 1999.
- [73] M. Mandjes. Pricing Strategies under Heterogeneous Service Requirements. In *Proceedings of IEEE INFOCOM*, 2003.
- [74] P. Marbach. Priority Pricing and Max-Min Fairness. In *Proceedings of IEEE INFOCOM*, 2002.
- [75] P. Marbach. Priority Service and Max-Min Fairness. *IEEE/ACM Transactions on Networking*, 11(5), 2003.
- [76] P. Marbach. Analysis of a Static Pricing Scheme for Priority Services. *IEEE/ACM Transactions on Networking*, 12(2), 2004.
- [77] P. Marbach and R. Berry. Downlink Resource Allocation and Pricing for Wireless Networks. In *Proceedings of INFOCOM*, 2002.

- [78] Y. Masuda and S. Whang. Dynamic Pricing for Network Service : Equilibrium and Stability. *Management Science*, 45(6) :857–869, 1999.
- [79] Lee McKnight and Joseph Bailey, editors. *Internet Economics*. MIT Press, 1997.
- [80] M. Mellia, R. Lo Cigno, and F. Neri. Measuring IP and TCP Behavior on Edge Nodes with Tstat. *Computer Networks*, 47, 2005.
- [81] H. Mendelson and S. Whang. Optimal Incentive-Compatible Priority Pricing for the m/m/1 Queue. *Operations Research*, 38(5) :870–883, 1990.
- [82] P. Merlin. *Le Transport Aérien*. Presse Universitaire Française, 2002.
- [83] D. Mills. Improved Algorithms for Synchronizing Computer Network Clocks. *IEEE/ACM Transactions on Networking*, 3(3), 1995.
- [84] H. Moulin and S. Shenker. Strategyproof Sharing of Submodular Costs : budget balance versus efficiency. Technical report, Rice University, 1997.
- [85] R. Núñez Quieja. *Processor-Sharing Models for Integrated-Services Networks*. PhD thesis, Eindhoven University of Technology, 2000.
- [86] A. Odlyzko. Paris Metro Pricing for the Internet. In *ACM Conference on Electronic Commerce (EC'99)*, pages 140–147, 1999.
- [87] A. Orda and N. Shimkin. Incentive Pricing in Multi-class Systems. *Telecommunication Systems*, 13, 2000.
- [88] A. Parekh and R. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks : The Multiple Node Case. *IEEE/ACM Transactions on Networking*, 2(2), 1994.
- [89] I.Ch. Paschalidis and J.N. Tsitsiklis. Congestion-Dependent Pricing of Network Services. *IEEE/ACM Transactions on Networking*, 8(2) :171–184, 2000.
- [90] V. Paxson. *Measurements and Analysis of End-to-End Internet Dynamics*. PhD thesis, University of California Berkeley, 1997.
- [91] M. Puterman. *Markov Decision Processes*. John Wiley & Sons, 1994.
- [92] P. Reichl, P. Flury, J. Gerke, and B. Stiller. How to Overcome the Feasibility Problem for Tariffing Internet Services : The Cumulus Pricing Scheme. In *IEEE International Conference on Communications*, 2001.
- [93] P. Reichl, D. Hausheer, and B. Stiller. The Cumulus Pricing model as an adaptive framework for feasible, efficient, and user-friendly tariffing of Internet services. *Computer Networks*, (43), 2003.
- [94] P. Reichl and B. Stiller. Edge Pricing in Space and Time : Theoretical and Practical Aspects of the Cumulus Pricing Scheme. In *Proceedings of ITC'01*, 2001.

- 
- [95] P. Reichl, B. Stiller, and S. Leinen. Pricing and Cost Recovery for Internet Services : Pratical Review, Classification, and Application of Relevant Models. *Netnomics*, 2(1), 2000.
- [96] J. Roberts. Internet Traffic, QoS and Pricing. *Proceedings of IEEE*, 92(9), 2004.
- [97] D. Ros and B Tuffin. A Mathematical Model of the Paris Metro Pricing Scheme for Charging Packet Networks. *Computer Networks*, 46(1), 2004.
- [98] D. Samet, Y. Tauman, and I. Zang. An Application of the Aumann-Shapley Prices for Cost Allocation in Transportation Problems. *Mathematics of Operations Research*, 9(1) :25–42, 1984.
- [99] A. Sampath, N. Mandayam, and J. Holtzman. Analysis of an Access Control Mechanism for Data Traffic in an Integrated Voice/Data Wireless CDMA System. In *Proceedings of the 46th Vehicular Technology Conference*, 1996.
- [100] D. Schmeidler and Y. Tauman. Incentive-Compatible Cost-Allocation Schemes. *Journal of Economic Theory*, 63 :189–207, 1994.
- [101] N. Semret. *Market Mechanisms for Network Resource Sharing*. PhD thesis, Columbia University, 1999.
- [102] S. Shenker, D. Clark, D. Estrin, and S. Herzog. Pricing in Computer Networks : Reshaping the Research Agenda. *Telecommunications Policy*, 20(3), 1996.
- [103] S. Sigit Puspito, W. Jarot, and M. Nakagawa. Transmission Power Control Techniques for the Reverse Link of OFDM-DS-SS-CDMA System. In *Proceedings of IEEE Symposium on Computers and Communications*, 1999.
- [104] V. Siris and C. Courcoubetis. Resource Control for Loss-sensitive Traffic in CDMA Networks. In *Proceedings of INFOCOM*, 2004.
- [105] D.J. Songhurst, editor. *Charging Communication Networks*. Elsevier, 1999.
- [106] S. Stidham. Pricing and Congestion Management in a Network with Heterogeneous Users. *submitted to IEEE Trans. Auto. Control*, 2002.
- [107] B. Stiller, J. Gerke, P. Reichl, and P. Flury. Management of Differentiated Services Usage by the Cumulus Pricing Scheme and a Generic Internet Charging System. In *IEEE/IFIP Symposium on Integrated Network Management*, 2001.
- [108] Y. Tauman. *The Aumann-Shapley prices : a survey*, chapter 18. Cambridge University Press, 1988.
- [109] B. Tuffin. Charging the Internet without Bandwidth Reservation : an Overview and Bibliography of Mathematical Approaches. *Journal of Information Science and Engineering*, 19(5), 2003.

- 
- [110] G. Van Kessel, R. Núñez Queija, and S. Borst. Differentiated Bandwidth Sharing with Disparate Flow Sizes. In *Proceedings of IEEE INFOCOM*, 2005.
  - [111] J. Van Mieghem. Price and Service Discrimination in Queuing Systems : Incentive Compatibility of  $Gc\mu$  Scheduling. *Management Science*, 46(9), 2000.
  - [112] M. Van Uitert. *Generalized Processor Sharing Queues*. PhD thesis, Center for Mathematics and Computer Science (CWI) Amsterdam, 2003.
  - [113] A. Viterbi. *CDMA Principles of Spread Spectrum Communication*. Addison-Wesley, 1995.
  - [114] Z. Wang. *Internet QoS : Architectures and Mechanisms for Quality of Service*. Morgan Kaufmann Publishers, 2001.
  - [115] H. Zhu and I. Chlamtac. An Analytical Model for IEEE 802.11e EDCF Differential Services. In *Proceedings of International Conference on Computer Communications and Networks*, 2003.