



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Institut National Polytechnique de Toulouse (INP Toulouse)*

Présentée et soutenue le 1/10/2015 par :

Jean-Baptiste DUPÉ

**Ordonnancement et gestion des ressources
pour un système de Télécommunications haut débit :
Optimisation de la bande passante satellite**

JURY

THOMAS NOËL	Professeur des Universités, Université de Strasbourg	Président du Jury
PASCALE MINET	Chargée de Recherche, INRIA	Rapporteur
TOUFIK AHMED	Professeur des Universités, IPB	Rapporteur
HARRY G. PERROS	Professeur des Universités, NC State University	Examineur
CÉDRIC BAUDOIN	Ingénieur de Recherche, Thales Alenia Space	Co-Encadrant
CAROLINE BÈS	Ingénieur de Recherche, CNES	Co-Encadrant
EMMANUEL CHAPUT	Maître de Conférences, INPT	Co-Encadrant
ANDRÉ-LUC BEYLOT	Professeur des Universités, INPT	Directeur

École doctorale et spécialité :

MITT : Domaine STIC : Réseaux, Télécoms, Systèmes et Architecture

Unité de Recherche :

Institut de Recherche en Informatique de Toulouse (IRIT)

Directeur(s) de Thèse :

André-Luc BEYLOT

Rapporteurs :

Pascale MINET et Toufik AHMED

C'est le temps que tu as perdu pour ta rose qui fait ta rose si importante.

Antoine de Saint-Exupéry, *Le Petit Prince*

Croire au soleil quand tombe l'eau.

Louis Aragon, *Le Fou d'Elsa*

REMERCIEMENTS

Voilà le seul contenu non scientifique de cette thèse, et pourtant il me semble que sans ces remerciements, elle ne serait jamais achevée. Si une thèse est parsemée de lourds moments de solitude, elle est aussi l'occasion d'échanges, de discussions et de moments partagés qui en font une expérience de travail d'une très grande richesse humaine, et qui fait bien vite partir le mauvais temps.

Je tiens à adresser, tout d'abord, mes plus sincères remerciements à M. André-Luc Beylot, mon directeur de thèse. Pour m'avoir donné la possibilité de faire cette thèse, pour m'avoir fait confiance, mais surtout pour m'avoir accompagné à chaque instant ; cette thèse n'aurait tout simplement pas vu le jour sans ce soutien. J'ai également bénéficié de l'encadrement remarquable, jour après jour, de M. Emmanuel Chaput. Je souhaite le remercier pour ses très nombreux conseils, sa patience (et il en a fallu !) et son enthousiasme pour mes recherches. La finesse de ses remarques m'a guidé tout au long de ma thèse, celle de ses plaisanteries m'a fait garder le sourire.

Cette thèse ayant une forte composante industrielle, j'ai pu m'appuyer sur l'ensemble de mes encadrants du CNES et de TAS pour orienter, affiner, et assurer mes travaux dans un contexte que je ne maîtrisais pas toujours. Merci à Mme Caroline Bès pour l'ensemble de ses conseils et remarques qui m'ont bien souvent remis en question au bon moment, et aussi pour la gentillesse qu'elle m'a témoignée. Merci également à Arnaud Deramecourt pour son soutien, pour ses remarques et son suivi attentif de ma thèse. Je dois enfin tous mes remerciements à M. Cédric Baudoin, pour les discussions que nous avons eues et pour les nombreux conseils qu'il m'a prodigué qui ont bien souvent été déterminants pour mes recherches. Il a su témoigner d'un intérêt constant pour mes recherches, travailler ensemble a été un vrai plaisir pour moi.

Je n'ai jamais manqué de rien pendant ces trois ans, ni de conseils, ni d'encouragements dans les moments difficiles, ni de confiance. Un grand merci à vous. Je garderai un souvenir impérissable de nos réunions d'avancement, surtout lorsqu'elles se terminaient de l'autre côté du canal !

Je dois également tous mes remerciements à Mme Corinne Mailhes, directrice du

TéSA. Son soutien aux instants clés de mes études a été déterminant, et la confiance qu'elle m'a témoignée m'a permis d'arriver jusqu'ici.

Je souhaite remercier mes rapporteurs, Mme Pascale Minet et M. Toufik Ahmed, pour leur relecture attentive de ma thèse, ainsi que pour leurs remarques pertinentes qui ont constitué un apport notable à ma thèse. Je remercie M. Thomas Noël d'avoir présidé mon jury de soutenance, ainsi que M. Harry Perros pour sa participation, malgré la distance.

Je voudrais également exprimer ma gratitude à l'ensemble de l'équipe IRT de l'IRIT, dans laquelle je me suis senti comme chez moi pendant ces trois ans. Merci pour les mémorables déjeuners du vendredi, que j'ai eu l'honneur d'organiser, et pour la gentillesse à mon égard. Une mention particulière pour les amateurs de la culture irlandaise, qui se reconnaîtront, et dont j'admire l'esprit combatif et le sens de l'humour. Même si cette fois, c'est "juste une"...

Merci également à toute la petite équipe des thésards, et en particulier Guillaume, Renaud, Patou, Nesrine, Aziz, Farouk, Mohammed, JGK, Éric et Élie. Merci pour tous ces bons moments passés ensemble, merci pour votre amitié et votre attention. Ces trois années sont passées bien vite en votre compagnie. Merci aussi à Mohammed, qui a partagé mon bureau (et quelques galères), pour sa bonne humeur et nos discussions si enrichissantes. Bon courage également à la relève, Dorin et Yoann.

Il me fallait également remercier ceux qui sont devenus, depuis quelques mois mes collègues : l'ensemble de l'équipe ST du CNES, et en particulier Caroline et Arnaud, avec qui j'ai grand plaisir à travailler à présent, et Thierry Robert, qui m'a accueilli au sein de son service.

J'ai également la chance d'avoir, à mes côtés un groupe, un clan sur qui j'ai pu compter pendant ma thèse, même si la moitié est encore persuadée que thésard, c'est quand même la belle vie (sans rancune, c'est vrai!). Merci aux "Loutres", et en particulier le trio de choc de l'outback australien : Xou, Bok et Charlie, qui adorent rouler en tout-terrain, ne craint personne au lever de coude, et a été là pour moi quand il le fallait. Merci aussi à Tysebab, Alex, Thomas, Aurélien et tous les autres pour les soirées, les fous rires, et le temps passé ensemble. Merci à Oriane, Mickaël, Toum et Marie de m'avoir supporté, pour les week-ends, les soirées démesurées à la coloc, pour les dîners et pour surtout votre gentillesse. Merci à Gaga, Laura et Elodie pour cette belle amitié qui dure. Merci aux colocs, Laura, Olivier, Amélie, Amélia et Jean-Fab, pour tout ce que nous avons partagé. Merci enfin à tous les "vieux" de l'n7, les amis du mauvais temps et d'ailleurs, trop nombreux pour être nommés ici, et surtout celle et ceux qui m'ont soutenu jusqu'au dernier instant, et qui se reconnaîtront. Merci pour tout !

Et à ceux qui m'ont toujours fait confiance, gardé dans les moments difficiles, entouré, donné l'essentiel pour que je réussisse : papa, maman, la famille Putaj et la famille Maillot, je souhaite dédier cette thèse. Votre fierté a fait la mienne, votre affection m'a porté.

Au moment de conclure, je me rends compte à quel point j'ai eu de la chance. Cette "école de la vie" qu'est la thèse vallait toutes les peines, tous les efforts que j'ai consentis pour arriver jusqu'ici. Les rencontres que j'ai pu faire y sont pour beaucoup. Je souhaite à chaque thésard de ressentir, au moment de refermer le manuscrit, ce même sentiment.

RÉSUMÉ

Les télécommunications par satellite ont connu ces dernières années un regain d'intérêt important, du fait de leur capacité à permettre la réduction de la fracture numérique. En effet, un satellite en orbite géostationnaire peut s'appuyer sur une très grande couverture et une capacité importante pour atteindre des zones où le déploiement des réseaux terrestres n'est pas envisageable, comme les transports (bateau, avion), ou bien les zones blanches, où il serait difficilement rentable. Traditionnellement concentrés sur la diffusion de télévision numérique, les dernières générations de standards reflètent cet engouement en faisant une place de choix à la transmission de données bidirectionnelle, notamment en permettant une prise en charge simple des protocoles de l'Internet. Le problème de l'ordonnancement dans ces systèmes devient alors particulièrement important, puisqu'il doit prendre en compte deux processus évoluant de manière totalement décorrélée. D'un côté, l'évolution de la demande des utilisateurs, dépendante des applications (vidéo, voix, données). De l'autre, l'évolution de la capacité du système, celle-ci étant tributaire des conditions de transmission : les fréquences utilisées dans ces systèmes sont particulièrement sensibles à l'atténuation due à l'eau dans l'atmosphère.

Cette thèse s'intéresse au problème de l'ordonnancement et de l'allocation de ressources, dans le but de fournir un service comparable aux réseaux terrestres en termes de services, en présentant les meilleures performances possibles. Si un certain nombre de propositions ont été faites sur le sujet, aucune ne prend en compte l'ensemble des contraintes d'un tel système. Outre le caractère variable de la capacité, la variabilité de la demande, conjuguée avec les contraintes de qualité de service constitue une difficulté supplémentaire. Enfin, il nous faut considérer la faisabilité de notre solution dans un contexte temps réel, nécessaire dans l'optique d'une implantation dans un système réel.

Nous avons ainsi développé une architecture d'ordonnanceur pour la voie Aller, reposant sur des fonctions d'utilité, permettant ainsi une formulation simple du compromis entre demande et capacité. Nous montrons comment cet algorithme pourrait être utilisable dans un système complet, à travers une implantation détaillée, de faible complexité, ainsi que des simulations de cas réels. Nous portons ensuite notre attention sur la voie Retour,

où nous proposons une méthode d'allocation de ressources prenant en compte de manière conjointe la qualité de service et la qualité du support pour délivrer une allocation à la fois conforme et performante. Les simulations montrent que notre algorithme obtient une efficacité et une meilleure gestion du trafic que des solutions de référence présentées dans la littérature.

ABSTRACT

Satellite telecommunications have seen a tremendous increase in interest, due to its ability to reduce the digital divide. In fact, a geostationary satellite can take advantage of its very wide coverage and high capacity to reach areas where deployment of a terrestrial network would not be possible, such as transports, or too expensive to be profitable, as in remote areas. Traditionally focused on digital television broadcasting, the latest generation of standards have evolved to reflect those new needs, dealing extensively with the transmission of interactive data, particularly by natively supporting Internet protocols. Scheduling has arisen as a major issue of those modern systems, since it has to deal with to highly uncorrelated processes : demand and capacity. Demand, on one side, evolves with user's needs, and therefore with the applications they are using : video, voice or data. Capacity, on the other side, depends on meteorological conditions over the satellite's cover, as the frequencies used in such systems are very sensitive to wet atmosphere attenuation.

This thesis aims to study the problem of scheduling and resource allocation, hoping to achieve a service that can match with terrestrial networks in terms of services, while showing the best possible performances. If numerous solutions were proposed on this topic, none is taking into account all of the current system's constraints. In addition to the variable nature of system's capacity, the conjunction of variable demand and quality of service constraints constitutes an additional issue. Furthermore, we have to consider the practicability of our solution in a real-time context, necessary if we aim for industrial use.

We have first developed a scheduler architecture for the Forward link, based on utility functions, thus allowing a simple formulation of the capacity versus demand compromise. We show, through a detailed low-complexity implementation and accurate simulations, how our algorithm could be used efficiently in an industrial context. We then focus on the Return link, where we propose a resource allocation method, taking into account quality of service and quality of transmission jointly to deliver an efficient yet consistent resource allocation. Simulations show that our algorithm achieves a better efficiency and

traffic handling than reference solutions presented in the literature.

TABLE DES MATIÈRES

Table des figures	xv
Liste des tableaux	xvii
Introduction	1
1 Étude des systèmes de Télécommunications par Satellite	7
1.1 Présentation du système	7
1.1.1 Composition du système	8
1.1.2 Conventions de nommage et fonctionnalités	10
1.2 Architecture Protocolaire	11
1.2.1 Architecture de la voie Aller	12
1.2.2 Architecture de la voie Retour	16
1.3 Demande et Capacité	19
1.3.1 Demande : Services et Utilisations	19
1.3.2 Capacité	22
1.4 Conclusions de l'étude	24
2 État de l'art des solutions d'ordonnancement et d'allocation de res- sources	27
2.1 Conventions de nommage et métriques	28
2.1.1 Définitions	28
2.1.2 Métriques	29
2.2 Outils analytiques d'ordonnancement	30
2.2.1 Critère Explicite	31
2.2.2 Théorie des jeux	34
2.2.3 Fonctions d'utilité	38
2.2.4 Conclusion de l'étude des outils analytiques	40

2.3	Revue des méthodes d'ordonnancement et d'allocation de ressources appliquées au Satellite	41
2.3.1	Voie Aller	41
2.3.2	La voie Retour	51
3	Algorithmes d'ordonnancement pour la voie Aller	57
3.1	Architecture de l'ordonnanceur	58
3.1.1	Choix de la Quality of Service (QoS)	58
3.1.2	Architecture proposée	59
3.2	Règles d'ordonnancement pour Digital Video Broadcasting - Satellite - Second Generation (DVB-S2)/Generic Stream Encapsulation (GSE)	61
3.2.1	Principes de l'adaptation des règles d'ordonnancement	61
3.3	Règles d'ordonnancement adaptées au satellite	64
3.3.1	Modified Latest Weighted Deadline First (M-LWDF) et Exponential Rule - PF (EXP-PF) : formulation	64
3.3.2	Algorithme utilisant les règles d'ordonnancement	66
3.4	Ordonnancement et fonctions d'utilité	67
3.4.1	Formulation du problème	67
3.4.2	Méthodes de résolution	68
3.4.3	Algorithme reposant sur le délai	70
3.5	Évaluation de performances	75
3.5.1	Scénarios	75
3.5.2	Scénario ciel clair	78
3.5.3	Scénario forte pluie	80
3.5.4	Résultats scénarios d'analyse	82
3.5.5	Complexité et discrétisation	83
3.5.6	Conclusion sur l'évaluation de performances	85
3.6	Conclusion sur l'ordonnancement voie Aller	86
4	Algorithmes d'allocation de ressources pour la voie Retour	89
4.1	Architecture du système	90
4.1.1	Paramètres et boucle d'allocation	90
4.1.2	Propriétés du trafic voie Retour	92
4.1.3	Algorithme de référence	93
4.2	Algorithme d'allocation de ressources optimisé	94
4.2.1	Formalisation	94
4.2.2	Problème d'optimisation	96

4.2.3	Implantation	99
4.3	Évaluation de performances	100
4.3.1	Scénario de référence	101
4.3.2	Résultats de simulation	102
4.3.3	Conclusion sur l'évaluation de performance	106
4.4	Conclusion sur l'allocation de ressources voie Retour	107
Conclusion		109
Bibliographie		115

TABLE DES FIGURES

1.1	Système de référence	8
1.2	Architecture de la voie Aller [13]	12
1.3	Boucle Adaptive Coding and Modulation (ACM) sur la voie Aller	14
1.4	Encapsulation GSE/DVB-S2	14
1.5	Hierarchie Multi-Frequency Time Division Multiple Access (MF-TDMA)	16
1.6	Allocation MF-TDMA	17
1.7	Relation entre conditions météo et débit pour un Terminal	23
2.1	Mesure discrète du temps	29
2.2	Architecture simple d'ordonnanceur	42
2.3	Architecture d'ordonnanceur proposée dans [7]	46
3.1	Architecture de l'ordonnanceur DVB-S2/GSE	60
3.2	Deficit Round Robin (DRR) adapté aux Base Band Frame (BBFRAME)	74
3.3	Fonctions d'utilité valorisant le temps d'attente	78
3.4	Temps d'attente moyen pour plusieurs algorithmes de référence	79
3.5	Temps d'attente moyen, scénario forte pluie	81
3.6	Résultats de simulation pour le scénario «iso-ModCod»	82
3.7	Résultats de simulation pour le scénario «iso-QoS»	82
3.8	Résultats de simulation pour le scénario «téléphonie»	83
3.9	Temps de calcul par BBFRAME	84
4.1	Chronogramme de la boucle d'allocation	91
4.2	Fonctions d'utilité pour la voie Retour	96
4.3	Algorithme d'allocation de ressources	100
4.4	Fonction de répartition empirique du temps d'attente	103
4.5	Résultats de simulation voie retour - délai moyen	104
4.6	Efficacité de l'allocation η_e	105
4.7	Charge système ρ_s et charge réelle ρ_r	105

4.8 Nombre d'itérations en fonction du temps 106

LISTE DES TABLEAUX

1.1	Caractéristiques pour plusieurs Modulation et le Codage (MODCOD) pour le DVB-S2	15
1.2	Exemple de répartition du trafic pour la voie Aller	22
1.3	Exemple de répartition des MODCOD de type ciel clair	24
3.1	Paramètres de QOS	75
3.2	distribution <i>Ciel clair</i>	76
3.3	distribution <i>Forte pluie</i>	77
3.4	Complexité d'algorithmes d'ordonnancement	84
4.1	distribution voie Retour – Ciel clair	101
4.2	Répartition des classes de QOS sur la voie Retour	102

ACRONYMES

- BBFRAME** Base Band Frame. xv, 13–15, 25, 28–30, 42–45, 48–50, 56–58, 60–63, 65–67, 69–75, 77, 84–86, 110
- DIFFSERV** Differentiated Services. 59, 70, 90, 92
- HoL** Head of Line. 33
- MODCOD** Modulation et le Codage. xvii, 13–15, 17, 22–24, 29, 30, 41–46, 49, 50, 52–55, 58–63, 65, 66, 68, 71–74, 76–78, 80–83, 86, 89–91, 93–96, 101–104, 111
- QoS** Quality of Service. xii, xv, xvii, 5, 12, 19–21, 25, 32, 34, 42–47, 49, 50, 52, 53, 55–61, 63, 64, 66, 67, 70, 73, 75–78, 80–83, 85, 86, 89–93, 95–97, 99, 100, 102–104, 107, 109, 110
- VoD** Video on Demand. 20, 92, 102
- VoIP** Voice over IP. 18, 20–22, 25, 34, 49, 53, 59, 101, 102
- WIMAX** Worldwide Interoperability for Microwave Access. 16
- ACM** Adaptive Coding and Modulation. xv, 13, 14, 17, 21, 22, 41, 57, 109
- ADSL** Asymmetric Digital Subscriber Line. 11, 19, 20
- AF** Assured Forwarding. 59, 65, 93, 103, 107
- ATM** Asynchronous Transfer Mode. 3, 16, 32
- BBS** BBFrame-by-BBFrame Scheduling. 69–71, 73, 75, 77, 79–82, 84–86, 110
- BE** Best Effort. 20–22, 53, 54, 59, 77, 78, 80, 82, 83, 93, 103, 104, 107
- BPS** BBFrame Periodic Scheduling. 70, 73, 75, 77–86, 110
- BTU** Burst-Time Unit. 17, 18, 55, 89–91, 93, 97, 99, 100, 102, 105, 106, 111
- CAC** Call Admission Control. 18, 53
- CBR** Constant Bit Rate. 44, 45, 53, 61, 95
- CCM** Constant Coding and Modulation. 13
- CRA** Constant Rate Assignment. 18, 19, 52, 90, 92

- CRDSA** Contention Resolution Diversity Slotted ALOHA. 18, 111
- DAMA** Demand Assigned Multiple Access. 18, 56, 90, 107, 111
- DRR** Deficit Round Robin. xv, 5, 32, 60, 73–75, 83–85, 87, 100, 104, 105, 107
- DVB-RCS2** Second Generation DVB Interactive Satellite System. 17, 18, 27, 29, 51–56, 89–92, 94, 102, 107, 109–111
- DVB-RCS** Digital Video Broadcasting - Return Channel via Satellite. 2, 16, 17, 51–55
- DVB-S2** Digital Video Broadcasting - Satellite - Second Generation. xii, xv, xvii, 2, 3, 12, 13, 15, 27, 41, 43, 49, 57, 59–61, 63, 76, 109
- DVB-S** Digital Video Broadcasting - Satellite. 1, 11, 12
- EDF** Earliest Deadline First. 34, 44
- EF** Expedited Forwarding. 59, 65, 75, 80, 83, 93, 102, 103, 107
- EXP-PF** Exponential Rule - PF. xii, 34, 61, 64, 65, 77, 79, 80
- FCA** Free Capacity Assignment. 52, 54, 89, 93, 96, 107
- FDMA** Frequency Division Multiple Access. 16
- FTP** File Transfer Protocol. 44, 45, 49, 54
- GSE** Generic Stream Encapsulation. xii, xv, 12, 13, 25, 43, 44, 49, 50, 56, 57, 60, 61, 63, 66, 73, 109
- HTTP** HyperText Transfer Protocol. 19, 21, 101
- IPTV** IP Television. 101
- ITU** International Telecommunication Union. 92
- LTE** Long Term Evolution. 16, 61, 92
- M-LWDF** Modified Latest Weighted Deadline First. xii, 34, 61, 64, 65, 72, 77, 79–81, 84–86
- MF-TDMA** Multi-Frequency Time Division Multiple Access. xv, 16, 17
- MPEG-2** Moving Picture Expert Group. 3, 12, 27
- MPEG2-TS** MPEG-2 Transport Stream. 12, 41
- MPE** Multiprotocol Encapsulation. 11–13, 25, 41
- NBS** Nash Bargaining Solution. 35–37, 55
- NOC** Network Operations Center. 9
- OFDMA** Orthogonal Frequency Division Multiple Access. 35, 61
- PDU** Protocol Data Unit. 17

- PF** Proportional Fairness. 33, 36, 38–40, 44, 45, 49, 50, 61–64, 72
- PHB** Per Hop Behaviour. 59, 70
- RBDC** Rate-Based Dynamic Capacity. 18, 19, 21, 52, 53, 90, 92, 93, 108
- RR** Round Robin. 3, 31, 32, 67, 73, 74
- SNR** Signal to Noise Ratio. 13
- SVC** Scalable Video Coding. 46–48
- TCP** Transmission Control Protocol. 92, 101, 103
- TDMA** Time Division Multiple Access. 16
- TDM** Time Division Multiplexing. 15
- TEB** Taux d’Erreur Binaire. 13, 14, 22, 23
- UBMT** Urgency-Based Maximum Throughput. 44, 78–85
- VBDC** Volume-Based Dynamic Capacity. 19, 21, 52, 53, 90, 92, 93, 108
- WFQ** Weighted Fair Queuing. 29, 32, 40
- WRR** Weighted Round Robin. 32, 41–43, 45, 47–49, 60, 78–80, 84, 85

INTRODUCTION

Peu de technologies peuvent se targuer d'être aussi populaires et universelles qu'Internet. Il est aujourd'hui un instrument nécessaire, sinon indispensable, pour un bon nombre d'utilisateurs, qu'ils soient particuliers, entreprises ou institutions. L'un de ses notables succès est son apparente ubiquité : Internet dispose à présent d'une très large couverture, bien au-delà du réseau téléphonique historique, grâce à l'utilisation des Radiofréquences, et plus particulièrement les technologies WiFi et mobiles. Elles permettent non seulement de diminuer largement les coûts de déploiement en comparaison d'un réseau filaire, mais offrent également une grande liberté d'utilisation, de déplacement, ce qui explique leur succès. Cependant, ce déploiement expansif se heurte parfois aux réalités du terrain, soit du fait d'un coût prohibitif par rapport au revenu attendu, comme dans les *zones blanches*, soit par l'impossibilité pure et simple d'utiliser les technologies terrestres, comme dans les transports aériens ou maritimes.

L'une des réponses à ce problème réside dans l'utilisation de satellites de télécommunications en orbite géostationnaire, qui étaient traditionnellement réservés à la diffusion de télévision numérique. L'orbite géostationnaire permet aux satellites d'offrir une très large couverture, du fait de leur position privilégiée, puisqu'il suffit de trois satellites, disposés à 120° d'écart chacun, pour couvrir 70% de la surface de la Terre, ce qui est considérable. De plus, la propagation du signal se fait en espace libre, permettant, malgré la distance, de très bonnes performances, et offrant un débit comparable – voire supérieur – aux accès terrestres. L'idée d'utiliser ces systèmes de diffusion pour des communications bidirectionnelles, tel l'accès à internet en zone blanche, semble donc particulièrement pertinente.

Cependant, les défis à relever avant d'offrir une solution crédible, tant en termes de performances que de coût et de complexité, sont nombreux. Les télécommunications par satellite ont longtemps profité de la popularité de la télévision numérique, qui s'est montrée remarquablement rentable. Aussi, les systèmes satellites ont-ils été conçus sur mesure pour la télévision, tel le Digital Video Broadcasting - Satellite (DVB-S) [1](1994),

et bien souvent sans prendre en compte la possibilité d'effectuer des communications bidirectionnelles, sinon de manière très marginale. Dans cette première génération de systèmes entièrement numériques¹, l'introduction *a posteriori* de la transmission de données interactive a été progressive, nécessitant de lourdes adaptations des standards existants, Digital Video Broadcasting - Return Channel via Satellite (DVB-RCS) [2](1999), et une complexité accrue en comparaison des systèmes terrestres. Afin de rendre les systèmes satellites compétitifs pour la transmission de données, la seconde génération de systèmes, Digital Video Broadcasting - Satellite - Second Generation (DVB-S2) [3](2003), a introduit de nombreuses améliorations et simplifications, permettant une prise en charge native et efficace des protocoles de l'Internet.

Pourtant, cette amélioration cache une réalité bien plus contrastée : si les standards modernes de télécommunications par satellite mettent à la disposition des ingénieurs des outils performants, leur utilisation se révèle parfois très complexe, et la réutilisation des architectures terrestres existantes n'est pas toujours possible.

Un effort de recherche a donc été entrepris afin de comprendre l'impact des nouveaux standards sur la performance et la complexité des systèmes de télécommunication. Cet effort s'est particulièrement manifesté en ce qui concerne l'ordonnement et l'allocation de ressources, du fait des profonds changements de la couche accès d'une génération à l'autre [4–9].

Ordonnement et allocation de ressources : du statique au dynamique

Deux processus définissent majoritairement l'ordonnement et l'allocation de ressources des systèmes de télécommunications par satellite. D'abord, intervient la dépendance de la qualité de la transmission vis-à-vis des conditions météorologiques, puisque les fréquences utilisées sont très sensibles à l'atténuation due à la vapeur d'eau. Le canal de propagation satellite est, de plus, bien différent des canaux mobiles, d'une part du fait qu'il ne présente que peu de multitrajet (vue directe), d'autre part par son évolution temporelle bien plus lente, essentiellement tributaire des phénomènes météorologiques. Cette dépendance influence directement le débit maximal dont un utilisateur peut disposer, et sa dynamique temporelle. Ensuite, le débit requis par le système afin d'acheminer les informations jusqu'aux utilisateurs, en d'autres termes la charge entrante du système, varie en fonction des applications et du nombre d'utilisateur actifs. Ce processus n'est pas, *a priori*, corrélé avec la variation des conditions météorologiques.

1. Des systèmes de diffusion analogique-numérique ont précédemment existé

Dans les systèmes de première génération, la dépendance à la qualité du lien est traitée de façon statistique : on peut en effet calculer, à l'aide de l'analyse des données météorologiques, un seuil de disponibilité pour le système. Ce seuil se définit comme la proportion du temps pendant laquelle un utilisateur peut recevoir un certain débit. Le dimensionnement du système se fait donc sur une métrique statistique *pire cas*, qui ne prend pas en compte le caractère dynamique des conditions de transmission. De plus, la diffusion de télévision engendre un débit constant, facilement prédictible. La vidéo utilise des trames Moving Picture Expert Group (MPEG-2), dont la longueur est fixée à 188 octets, ce qui permet de calculer, en fonction du codec vidéo utilisé, le débit nécessaire. Une fois le nombre de chaînes et les codecs définis, il est possible de calculer explicitement le débit demandé pour acheminer l'ensemble de ces chaînes. Ainsi, les systèmes de télécommunications de première génération se caractérisent par leur nature statique, voire périodique, puisque le système possède une capacité et une demande qui n'évoluent pas avec le temps.

Lors de l'introduction de la transmission de données, les algorithmes d'ordonnancement ont été adaptés à ce contexte, et ont conservé l'architecture des systèmes de première génération. S'appuyant sur des algorithmes précédemment développés sur des systèmes terrestres disposant de caractéristiques similaires, comme Asynchronous Transfer Mode (ATM), les ordonnanceurs de première génération se caractérisent par un comportement très statique, où le paramétrage s'effectue en fonction d'hypothèses connues sur la nature du trafic, comme le débit, la tolérance à la gigue ou aux pertes. Des algorithmes simples, dérivés du Round Robin (RR), ont été largement utilisés dans ce contexte [10].

Mais l'introduction de la seconde génération de systèmes, résolument tournée vers le monde IP (comme DVB-S2 [3]) a rendu ces considérations en partie obsolètes. La gestion du trafic de type Internet suppose un traitement asynchrone, des paquets de longueur variable, et une demande évoluant en permanence. Dans la couche physique, l'utilisation de méthodes d'adaptation dynamique à la qualité de la transmission a rendu les systèmes certes plus performants, mais suppose une capacité dynamique, elle aussi évoluant en permanence. Par conséquent, il est crucial pour les algorithmes d'ordonnancement de prendre en compte dès la conception ce caractère dynamique. L'introduction d'une voie Retour plus performante suppose également des algorithmes d'allocation de ressources efficaces, qui sont, pour la voie Retour, le pendant des algorithmes d'ordonnancement sur la voie Aller.

Contributions et Plan

Dans cette thèse, nous montrons en quoi les propositions actuelles d'ordonnanceurs pour les systèmes de télécommunications par satellite ne permettent pas d'utiliser au mieux les ressources importantes des systèmes modernes de télécommunications par satellite. L'intégration de trafic très différents, conjuguée au besoin de flexibilité des algorithmes rend la conception d'algorithmes particulièrement complexe. Aussi, avons-nous choisi de nous concentrer sur la conception d'outils mathématiques, permettant d'analyser clairement les étapes de cette conception, afin d'en étudier la pertinence et l'influence.

Au regard des résultats obtenus en appliquant la méthode développée, nous pouvons résumer nos contributions de la manière suivante :

- Nous avons proposé une classification exhaustive des algorithmes d'allocation de ressources et d'ordonnement, en insistant sur l'ensemble de la conception de l'algorithme : du choix de la méthode d'évaluation, à l'implantation effective. Nous avons relevé ainsi un manque de cohérence quant à l'implantation dans nombre de travaux, et proposons une conception couvrant l'ensemble de la conception de l'algorithme.
- Prenant appui sur la méthode décrite, nous avons proposé un ordonnanceur performant et flexible, qui réponde aux attentes d'un système de dernière génération. Nous avons, à l'aide d'outils d'optimisation convexe, caractérisé un type d'ordonnement optimal et déduit une méthode de résolution rapide. Nous avons sélectionné la théorie de l'utilité, dont John Von Neumann fit usage dans ses travaux de référence sur la théorie des jeux [11]. Les fonctions d'utilité fournissent des outils de modélisation et d'optimisation, et nous justifierons leur utilisation dans les chapitres suivants. De part ses performances, cet ordonnanceur présente une utilisation efficace et dynamique des ressources du système, tout en offrant les garanties nécessaires à des trafics fortement contraints. De plus, l'implantation réalisée possède une complexité suffisamment faible pour qu'un développement puisse désormais être envisagé sur des systèmes réels.
- En ce qui concerne l'allocation de ressources sur la voie Retour satellite, nous avons appliqué notre méthode de conception afin d'obtenir un algorithme offrant des performances supérieures à l'état de l'art actuel dans le domaine. Les outils que nous avons développés pour le satellite nous permettent une analyse rigoureuse du processus d'allocation de ressources. Nous avons également pu observer de façon claire les points-clés qui régissent la performance de ces algorithmes. Nous

proposons plusieurs pistes d'améliorations pour de futurs systèmes.

Notre analyse de l'ordonnancement et de l'allocation de ressources sera composée de quatre parties, dont le détail est donné ci-après.

Chapitre 1 : Ce chapitre présente l'architecture des Systèmes de Télécommunications par Satellite de dernière génération. Après une présentation des différents éléments constitutifs du système, nous abordons successivement la voie Aller et la voie Retour, sous leurs aspects fonctionnels et protocolaires. Nous montrons dans quelle mesure les protocoles utilisés, ainsi que le système lui-même, peuvent influencer l'ordonnancement sur la voie Aller, et l'allocation de ressources sur la voie Retour. Enfin, une description des processus de demande et de capacité est proposée, permettant l'introduction aux problèmes d'ordonnancement et d'allocation de ressources.

Chapitre 2 : Dans ce chapitre, nous examinons d'abord les outils analytiques à notre disposition permettant de modéliser notre problème. Cette étude nous permet d'établir les fondements mathématiques utilisés dans nombre de travaux dans le domaine. Nous proposons ensuite une analyse des méthodes proposées spécifiquement pour les télécommunications par satellite, pour la voie Aller puis pour la voie Retour, et montrons en quoi elles ne permettent pas de répondre de manière efficace à notre problème.

Chapitre 3 : Dans ce chapitre, nous étudions les ordonnanceurs sur la voie Aller des systèmes satellites. Nous présentons un ordonnanceur, permettant une prise en compte du caractère dynamique du système, et fondé sur les fonctions d'utilité. Outre les avantages inhérents à cette méthode, nous présentons également une implantation de cet algorithme à l'aide d'un ordonnanceur classique de type Deficit Round Robin (DRR).

Chapitre 4 : La voie Retour est traitée dans ce dernier chapitre, en faisant usage de la méthode de conception mise en place au Chapitre 3. Une attention particulière est portée sur les paramètres ayant la plus forte influence sur la performance globale du système, particulièrement le traitement de la Quality of Service (QoS), ainsi que l'équité. Reposant sur un problème d'optimisation, nous décrivons également une méthode de résolution, ainsi qu'une évaluation de ses performances.

ÉTUDE DES SYSTÈMES DE TÉLÉCOMMUNICATIONS PAR SATELLITE

Dans cette première partie, nous cherchons à comprendre les contraintes et les besoins des Systèmes de Télécommunication par Satellite, concernant l'ordonnancement et l'allocation de ressources. Après une définition du cadre fonctionnel et protocolaire dans lequel nous allons travailler, deux analyses sont successivement conduites, l'une sur les services offerts, et la demande des utilisateurs, l'autre sur la capacité offerte par de tels systèmes. Il ressort de cette étude un problème majeur pour l'ordonnancement et l'allocation de ressources, dû au caractère dynamique et décorrélé de la demande et de la capacité du système.

1.1 Présentation du système

Bien que semblables d'un point de vue fonctionnel à leurs équivalents terrestres, les Systèmes de Télécommunications par Satellite présentent des caractéristiques très disparates selon les applications, bien loin de la relative homogénéité des réseaux sans fil terrestres. En effet, le terme de Télécommunications par Satellite regroupe aussi bien les réseaux de collecte de données (tel ARGOS) que la navigation (GPS, GALILEO), ou encore ceux réservés principalement à la téléphonie (Inmarsat, Iridium). L'application principale, que nous considérons dans cette étude, est l'accès *broadband*, qui permet un accès à Internet et à d'autres services IP par satellite. On peut encore citer les systèmes de *backhaul* où l'objectif est de relier deux sites distants, appartenant au cœur du réseau. Ces liens présentent généralement un débit élevé, et justifient donc l'utilisation de ressources dédiées, où l'ordonnancement est absent – sinon élémentaire.

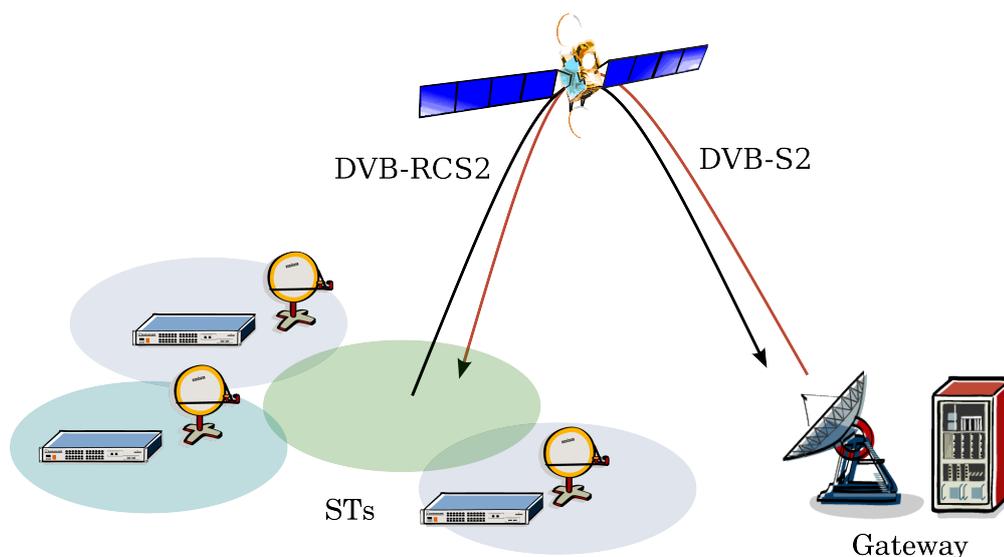


FIGURE 1.1 Système de référence

Notre objectif ici est de proposer un système broadband permettant d'offrir un complément aux systèmes sans fil terrestres d'accès à Internet, ainsi qu'une gamme similaire de services, dans les zones où le déploiement d'un réseau terrestre entraîne des coûts prohibitifs, voire n'est tout simplement pas envisageable. Un tel système pourrait jouer un rôle important dans la réduction de la fracture numérique, en proposant un service adapté à des utilisateurs résidentiels dans les zones blanches, ainsi que dans les transports. Il n'exclut pas pour autant d'autres applications citées plus haut, notamment le *backhauling*, mais nous considérons ces dernières comme marginales comparativement au service d'accès broadband, et ayant par conséquent une influence limitée sur le système. Nous présentons ici un système de référence répondant à ce besoin, afin de clarifier le cadre dans lequel notre étude va se dérouler.

1.1.1 Composition du système

Le système d'accès broadband de référence, présenté sur la figure 1.1, consiste en un certain nombre de Terminaux Satellites (ST), interconnectés, *via* un satellite en orbite géostationnaire, à une Passerelle (*Gateway*). Les terminaux sont situés chez les abonnés et permettent d'interconnecter un réseau domestique, constitué généralement de quelques machines (ordinateurs, téléphones, tablettes). La Passerelle est elle située chez l'opérateur, interconnectée avec le cœur de réseau. On peut résumer les fonctionnalités de chaque élément du système de la façon suivante.

La Passerelle

La Passerelle (ou *Gateway*) est l'élément principal du système, interconnectant le réseau d'accès satellite à un réseau terrestre de grande capacité. Elle est responsable de la gestion des ressources radio, en émission et en réception, ainsi que de la collecte du trafic à destination et provenant des Terminaux. Bien que les fonctionnalités soient généralement partagées entre la Gateway (modem), le Network Operations Center (NOC) (plan de Contrôle et Supervision) et le Hub (interconnexion), nous ne concentrons notre étude que sur une partie de ces fonctionnalités, qui ne nécessite pas de distinction particulière entre Hub, Passerelle et NOC. Le terme générique de Passerelle sera donc utilisé pour désigner les trois fonctions.

Les Terminaux

Les Terminaux Satellite sont les équipements les plus nombreux du système, entre quelques milliers pour les systèmes actuels, et plusieurs dizaines de milliers pour les systèmes futurs. Ils permettent d'interconnecter le réseau domestique avec le réseau d'accès Satellite. Leur fonctionnement est régi par la Passerelle, et ils ne prennent que peu d'initiatives dans la gestion des ressources radio, sur un principe de relation très asymétrique similaire à celui que l'on trouve dans les réseaux terrestres.

Ils sont caractérisés par des ressources (mémoire, CPU) faibles, et une puissance limitée, de façon à assurer un coût le plus bas possible. De fait, ils sont considérés comme un des facteurs limitants : il est difficile de leur faire effectuer des tâches nécessitant d'importantes ressources de calcul. La ressource radio réservée par Terminal, étant faible, est un autre facteur limitant la capacité. Si le débit reçu par les Terminaux peut être important, le débit engendré par ces derniers reste faible, correspondant à quelques machines utilisées simultanément dans le réseau domestique, ou bien dans un petit réseau d'entreprise.

Étant donnée la topologie adoptée par ce réseau, aucune différence particulière n'est faite entre les Terminaux fixes, nomades ou mobiles. En effet, la gestion de la mobilité (routage, signalisation), n'a pas un impact suffisant sur l'ordonnancement ou l'allocation de ressources pour justifier un traitement séparé. De plus, la taille importante des spots dans un tel système réduit de façon importante le nombre de handovers, en comparaison d'un réseau mobile terrestre.

Le Satellite

Un satellite en orbite géostationnaire permet la communication entre la Passerelle et les Terminaux, selon une topologie en étoile, où les communications se font en point à point. Nous limiterons ici notre étude au cas d'un satellite transparent, c'est-à-dire faisant office de répéteur de niveau physique, et ayant un impact négligeable du point de vue de la couche accès. Les communications se font donc exclusivement entre la Passerelle et les Terminaux, même dans le cas d'une communication entre deux Terminaux. Ce choix est justifié par le coût et la complexité des solutions actuelles avec une charge utile régénérative, qui permet de placer une couche accès à bord du satellite afin de faire communiquer directement les terminaux entre eux, réduisant de moitié le délai. De plus, pour l'accès à Internet et aux autres services, les communications entre Terminaux sont marginales, le contenu recherché par les utilisateurs étant souvent situé derrière la Passerelle.

1.1.2 Conventions de nommage et fonctionnalités

Le contexte historique de diffusion de télévision est très présent dans les Systèmes de Télécommunication par Satellite, et cette influence se ressent sur les conventions de nommage adoptées pour les communications Terminal-Passerelle et Passerelle-Terminal. Elle est également présente dans les différences importantes qui existent entre les architectures protocolaires de ces deux sens de communication.

La voie Aller

On parle de voie Aller pour la communication de la Passerelle aux Terminaux Satellite. Cette voie est utilisée traditionnellement pour la diffusion de télévision, mais a été adaptée pour les communication en *unicast*, à destination d'un Terminal en particulier. Si le support hertzien est un atout pour la diffusion, il suppose un mode de fonctionnement différent en unicast, où chaque terminal recevra l'ensemble des données, et devra donc être en mesure de déterminer celles qui lui sont destinées, par un mécanisme d'adressage. La tâche prépondérante de la Passerelle sera de déterminer, à chaque instant, le Terminal vers lequel acheminer les données qu'elle doit envoyer, et les critères selon lesquels choisir ces données. Ce problème est appelé ordonnancement dans la suite de cette étude. La performance des mécanismes d'ordonnancement est primordiale dans un tel système : étant donné le volume très important de données à traiter, une différence minime de performances aura un impact très important sur l'ensemble du système. L'optimisation des mécanismes d'ordonnancement présente donc un grand intérêt, particulièrement si

elle permet d'améliorer le rapport entre coût et performances. Dans cette thèse, nous ne retiendrons donc que les mécanismes de la couche accès de cette voie Aller, là où est implanté l'ordonnement.

La voie Retour

On parle de voie Retour pour la communication des Terminaux Satellite vers la Passerelle. Cette voie a été introduite afin de permettre des communications bidirectionnelles par satellite. Contrairement à certains systèmes, où une voie Retour terrestre était envisagée, nous considérons une voie Retour exclusivement par Satellite. Outre les avantages de déploiement, cette solution reste plus simple en termes de signalisation, d'architecture (routage), et présente un meilleur débit qu'une voie Terrestre fortement atténuée (comme la paire de cuivre des réseaux Asymmetric Digital Subscriber Line (ADSL)).

Du fait du partage de la voie Retour entre tous les Terminaux, pour communiquer avec la Passerelle, il est nécessaire de diviser efficacement la ressource radio allouée au système entre les Terminaux. L'objectif principal de la voie Retour est donc de diviser la ressource, et de la répartir de manière à satisfaire les besoins des Terminaux et ainsi des services utilisés. On distinguera ici deux processus : le premier est la méthode de multiplexage, ou d'accès multiple utilisée, qui consiste à spécifier comment les ressources spectrales sont divisées, le deuxième est la méthode d'allocation, ou comment les ressources sont attribuées aux Terminaux. Si la méthode de multiplexage est généralement normalisée, ce n'est pas le cas des algorithmes d'allocation de ressources. Ainsi, dans l'objectif d'optimiser conjointement le coût et les performances de notre système, l'allocation de ressources constitue un sujet d'étude particulièrement intéressant. La méthode d'accès (comprenant les deux processus) devra s'adapter dynamiquement aux besoins des Terminaux, ainsi qu'aux contraintes de transmission, et il est aisé de voir que sa performance est capitale pour le système dans son ensemble. Notre étude se portera donc, dans le cas de la voie Retour, sur ces algorithmes d'allocation de ressources.

1.2 Architecture Protocolaire

Dans cette partie, nous analysons les architectures protocolaires de référence pour la voie Aller ainsi que pour la voie Retour, et leur impact respectif sur l'ordonnement et l'allocation de ressources. C'est, en effet, à travers ces mécanismes protocolaires que se déroulera leur mise en place effective.

Les systèmes numériques de télécommunication par satellite ont connu deux générations majeures. La première, composée de DVB-S [1] et de Multiprotocol Encapsulation (MPE),

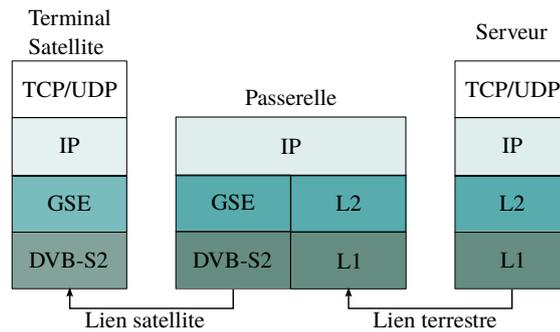


FIGURE 1.2 Architecture de la voie Aller [13]

et la seconde, constituée de DVB-S2 [3] et Generic Stream Encapsulation (GSE) [12]. Si ces deux générations permettent la diffusion de télévision numérique de manière similaire, la prise en charge des protocoles autres que ceux de MPEG-2 diffère d'une génération à l'autre.

Plus précisément, là où MPE était une adaptation aux protocoles du monde de l'Internet, GSE offre une prise en charge dès la conception, permettant une gestion simplifiée des protocoles tels qu'IP, MPLS, Ethernet.

On remarque cependant de nombreuses similitudes dans l'architecture des solutions de première et de deuxième génération. On pourrait alors tirer parti des travaux existants, conçus pour DVB-S, afin d'adapter les mécanismes, mais il faut s'assurer au préalable que les évolutions introduites par DVB-S2 ne remettent pas en cause le fonctionnement des mécanismes originaux.

1.2.1 Architecture de la voie Aller

L'architecture protocolaire de la voie Aller repose sur deux protocoles, GSE et DVB-S2. La figure 1.2 présente cette architecture.

Le protocole GSE [12] fait office de couche d'adaptation, permettant la prise en charge d'un très grand nombre de protocoles (IP, Ethernet, MPLS, etc...) en introduisant la notion de «flux» GSE, qui peut correspondre à un tunnel de niveau 2 ayant ses caractéristiques spécifiques de Qualité de Service (ou QoS). Ce protocole permet également une compatibilité avec les flux de type MPEG-2, utilisés pour la diffusion de télévision, qui peuvent être transportés avec MPEG-2 Transport Stream (MPEG2-TS) ou GSE. Une des principales caractéristiques de GSE est la prise en compte de paquets de taille variable, contrairement à la précédente génération, reposant sur MPE, qui demandait une segmentation en trames MPEG2-TS de 188 octets. Cet avantage est décisif, puisqu'il permet une gestion bien plus simple des protocoles de type IP : l'overhead est réduit

d'un facteur 2 par rapport à MPE [12], et le traitement nécessaire à la gestion des paquets de taille variable est simplifié.

Le protocole DVB-S2 définit le format des trames envoyées sur le lien satellite. Ce format est très éloigné des technologies terrestres. Dans le but de garantir un Taux d'Erreur Binaire (TEB) très bas (10^{-7}) les trames DVB-S2, appelées Base Band Frame (BBFRAME) ont une taille importante, fixée après codage à 64800 (trames longues) ou 16200 bits (trames courtes). La taille importante des BBFRAME permet l'utilisation efficace de méthodes de codage avancées, telles que les Turbocodes [14] ou encore les codes LDPC [15] dans DVB-S2. Une BBFRAME est définie par la Modulation et le Codage (MODCOD) qu'elle utilise. Cette flexibilité permet d'adapter la robustesse de la transmission à la qualité du lien, afin de fournir un TEB indépendant de cette qualité. Le schéma d'encapsulation de GSE dans DVB-S2 est présenté sur la figure 1.4. On distingue deux étapes dans la formation de la trame DVB-S2 : le codage et la modulation. On utilisera dans la suite le terme de BBFRAME pour désigner les trames envoyées, mais cette appellation est théoriquement dévolue à l'étape de codage, où une partie variable de la trame est dédiée à la redondance. L'étape de modulation correspond à la PLFRAME : comme la BBFRAME a une taille fixe après codage, la modulation va donner le temps nécessaire à la transmission de cette trame. En combinant ces deux étapes, on peut définir le débit utile d'une BBFRAME comme la quantité de données utiles envoyées, dépendante du taux de codage, divisée par le temps de transmission, dépendant de la modulation.

Une méthode, nommée Adaptive Coding and Modulation (ACM), permet d'adapter dynamiquement le MODCOD requis par un terminal en fonction de ses conditions de transmission. Cette technologie suppose l'utilisation d'une voie Retour satellite, où l'envoi périodique de signaux connus (ou pilotes) sur la voie Aller va permettre d'estimer la qualité de la liaison, généralement, mesurée en termes de rapport signal sur bruit (en décibels), ou Signal to Noise Ratio (SNR). Le Terminal sera alors en mesure d'envoyer son MODCOD optimal à la Passerelle, qui pourra alors envoyer les informations destinées au Terminal en utilisant ce MODCOD. Cette technique permet une utilisation optimale de la bande passante, contrairement aux méthodes à MODCOD fixe de type pire cas. Elle est donc largement plébiscitée pour augmenter la capacité des systèmes, mais présuppose de prendre en compte le caractère dynamique de cette même capacité, dû aux évolutions météorologiques sur la couverture du satellite. Ceci constitue une évolution majeure au regard des systèmes de diffusion de télévision, dont le fonctionnement peut être considéré comme régulier et invariant dans le temps, reposant sur des mécanismes à modulation et codage constants (Constant Coding and Modulation (CCM)). On parlera de *boucle* ACM pour désigner le mécanisme par lequel le Terminal informe périodiquement de la

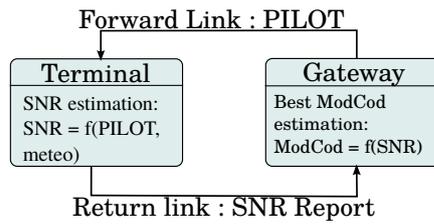


FIGURE 1.3 Boucle ACM sur la voie Aller

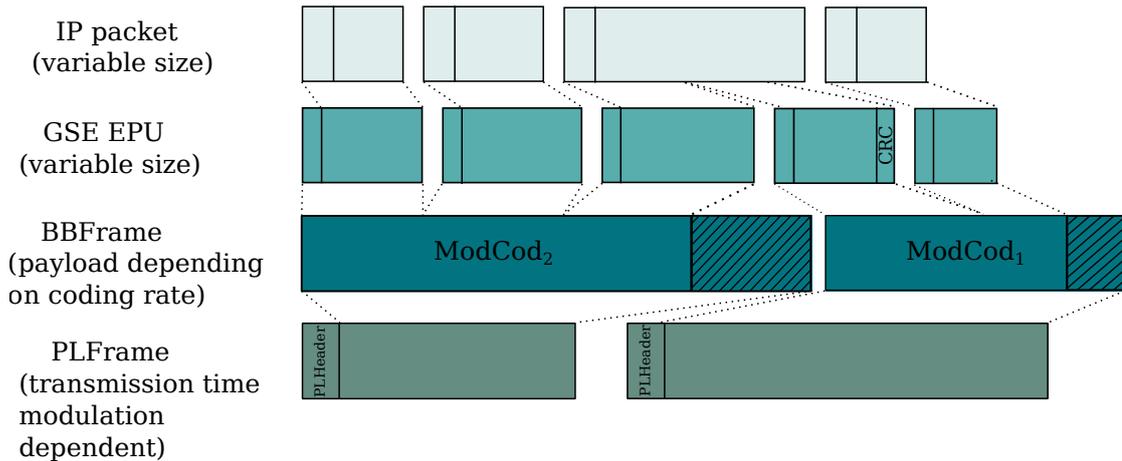


FIGURE 1.4 Encapsulation GSE/DVB-S2

qualité de ses transmissions, et l'impact de cette information sur le MODCOD attribué au Terminal par la Passerelle. Ce comportement commande-réponse est caractéristique d'un système bouclé, comme présenté sur la figure 1.3.

Une conséquence importante de cette technique est que la charge utile d'une BBFRAME dépend du MODCOD : à taille fixe, plus la redondance introduite avec le codage est importante, moins il y a de place disponible pour les données. De plus, le temps d'émission de la trame dépendra de l'ordre de la modulation, à débit symbole constant, comme indiqué sur la figure 1.4.

Une BBFRAME disposant d'un taux de codage faible et d'une modulation d'ordre faible (ou plus simplement ayant un MODCOD faible) sera adaptée à des communications perturbées, garantissant un débit certes faible, mais à TEB fixé. De même, une BBFRAME ayant un fort taux de codage (proportion de données utiles élevée), et une modulation d'ordre élevé (fort MODCOD, ayant un nombre de bits par symbole important) sera adaptée à de bonnes conditions de transmission, offrant un débit utile plus élevé.

Deux remarques importantes peuvent être faites. Tout d'abord, étant donnée la taille importante des BBFRAMES en comparaison des paquets IP, plusieurs de ces derniers

peuvent être encapsulés puis concaténés dans une BBFRAME, ce qui aura un impact important sur l'ordonnancement. En effet, si plusieurs paquets IP sont encapsulés dans une BBFRAME, il est nécessaire que les Terminaux auxquels ils sont destinés puissent décoder cette BBFRAME, et qu'ils disposent par conséquent d'un lien dont la qualité est suffisante. Cela pose une condition sur les paquets qui peuvent être encapsulés dans une BBFRAME, et par conséquent sur les Terminaux qui pourront être servis. De plus, il semble important de noter que charge utile (volume de données transportables) et temps de transmission d'une BBFRAME peuvent grandement varier, comme l'indique la table 1.1. Cet exemple est donné pour une PLFrame de 64800 bits et un débit symbole de 10Mbauds. Le choix de la BBFRAME aura donc non seulement une influence sur les paquets qui peuvent y être encapsulés, mais également sur le volume disponible pour ces paquets, ainsi que sur leur temps de transmission, et par conséquent sur le débit utile.

Il est possible d'utiliser une BBFRAME ayant un faible MODCOD pour encapsuler un paquet destiné à un Terminal capable de décoder un MODCOD plus élevé. Nous appelons cette technique le *déclassement*. Elle est utilisée lorsqu'une BBFRAME ne peut être remplie avec des paquets de son MODCOD, et qu'il reste des paquets à envoyer ayant un MODCOD plus fort, qui permettent alors de remplir la BBFRAME et de maximiser l'efficacité spectrale.

Modulation	Codage	Charge utile (bits)	Temps d'émission (ms)	Débit
BPSK	1/2	32400	6.48	5 Mbits/s
QPSK	1/3	21600	3.24	6.67 Mbits/s
8-PSK	2/3	43200	2.16	20 Mbits/s
16-APSK	5/6	54000	1.62	33.3 Mbits/s

TABLE 1.1 Caractéristiques pour plusieurs MODCOD pour le DVB-S2

La Passerelle dispose, pour la transmission sur la voie Aller, d'un certain nombre de porteuses (bande de fréquences) fixées par l'opérateur. L'émission d'une BBFRAME occupe la totalité d'une porteuse. Dans cette étude, nous nous limitons à l'étude d'un seul Time Division Multiplexing (TDM) (ou une seule porteuse), soit un ordonnancement BBFRAME par BBFRAME. Les Terminaux pour l'accès broadband ne permettent généralement de recevoir qu'une seule porteuse à la fois : il n'est donc pas possible, en tout état de cause, d'exploiter la diffusion sur plusieurs porteuses vers le même Terminal dans l'accès de type broadband considéré ici.

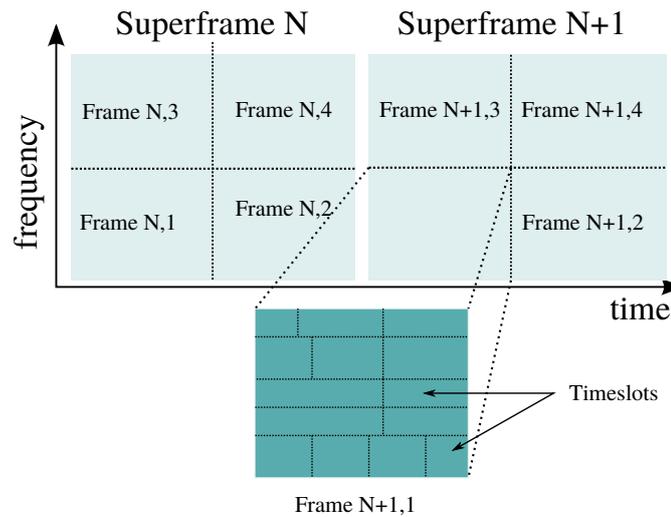


FIGURE 1.5 Hiérarchie MF-TDMA

1.2.2 Architecture de la voie Retour

MF-TDMA

Dans un Système de Télécommunications par Satellite, une partie de la bande passante est réservée aux communications sur la voie Retour. Il est alors nécessaire de diviser cette bande passante entre les Terminaux souhaitant transmettre des données. La voie Retour Satellite est caractérisée par son mode d'accès : le Multi-Frequency Time Division Multiple Access (MF-TDMA).

Le principe de MF-TDMA est de diviser cette ressource temporellement et fréquentiellement, afin de fournir à chaque Terminal une unité temps-fréquence dans laquelle il sera le seul à transmettre. Cette méthode combine les méthodes d'accès Time Division Multiple Access (TDMA) et Frequency Division Multiple Access (FDMA), afin d'optimiser l'utilisation des ressources en augmentant la granularité de l'allocation.

L'allocation de ressources consistera alors à déterminer les unités allouées à un Terminal donné. Deux niveaux hiérarchiques permettent de subdiviser l'espace temps-fréquence : la Trame (*frame*), composée d'unités temps fréquence, et la Supertrame (*Superframe*), composée de Trames, comme indiqué sur la figure 1.5. On retrouve cette structure dans la plupart des réseaux terrestres sans fil à accès multiple, tels que les réseaux mobiles, comme Long Term Evolution (LTE) et Worldwide Interoperability for Microwave Access (WiMAX).

Le précédent standard [2], DVB-RCS, décrivait une voie Retour fondée sur l'ATM et l'AAL 5. Cette solution permettait d'obtenir, à partir de paquets de taille variable, des

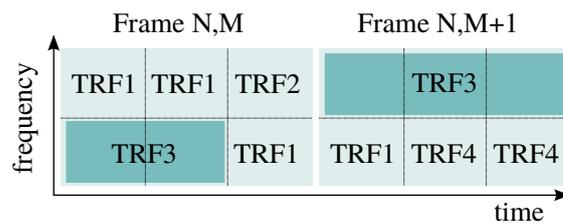


FIGURE 1.6 Allocation MF-TDMA

cellules de taille constante, plus simples à ordonnancer, mais au prix d'une complexité protocolaire et d'un *overhead* plus importants. Ces cellules étaient ensuite transmises avec le MODCOD adapté à la transmission. De ce fait, la taille d'un élément temps-fréquence, appelé *timeslot*, était dépendante du MODCOD, comme représenté sur la figure 1.5, pour la Supertrame N+1. Dans les faits, on réservait une partie des ressources pour chaque type de MODCOD, de manière à pouvoir avoir un format de Trame fixe. Les timeslots ainsi créés étaient ensuite alloués aux Terminaux selon leurs besoins. Ce mode de fonctionnement, où le format de la trame était statique, supposait de calculer la proportion des ressources qu'il fallait allouer à chaque MODCOD, et ce en fonction de la demande et des conditions de transmission des Terminaux. Avec un système dynamique, utilisant intensivement l'ACM, ce mode de fonctionnement ne semble plus adapté.

Dans la deuxième génération du standard [16] Second Generation DVB Interactive Satellite System (DVB-RCS2), l'unité d'allocation est la Burst-Time Unit (BTU), dont la durée et la bande sont définies et indépendantes du MODCOD. DVB-RCS2 comporte également une boucle ACM, dont le fonctionnement est comparable à celui de la voie Aller. Du fait de l'utilisation de codages et de modulations différentes, un Terminal demandera plus ou moins de BTU, selon son MODCOD, pour transmettre un même volume de données. Cette allocation indépendante des MODCOD permet une flexibilité accrue en comparaison de DVB-RCS, bien adaptée aux systèmes dynamiques. De fait, dans DVB-RCS2, il est possible de changer la composition des trames, et par conséquent l'allocation des BTU, entre la Supertrame N et la Supertrame N+1.

Un exemple d'allocation est présenté figure 1.6, avec des Terminaux (timeslots de trafic TRF1, 2, 3 et 4), ayant un MODCOD différent. On appelle timeslot dans DVB-RCS2 l'ensemble des BTU nécessaires à la transmission d'une Protocol Data Unit (PDU) sur la voie Retour. Dans l'exemple de la figure 1.6, là où le Terminal 1 peut transmettre une PDU dans un timeslot, le Terminal 3 en requiert trois.

DAMA et estimation des requêtes

La méthode d'allocation utilisée ici, où les ressources sont partagées entre les Terminaux selon leur demandes, est appelée Demand Assigned Multiple Access (DAMA). Chaque Terminal se voit allouer des ressources qu'il est le seul à pouvoir utiliser (sous forme de BTU, ici), en fonction de sa demande (ou requête). Il est possible, à chaque instant, de donner la composition de la Trame, ainsi que le Terminal auquel est associé une BTU donnée.

Cette méthode est dite déterministe, par opposition aux méthodes à contention de type Slotted ALOHA et Contention Resolution Diversity Slotted ALOHA (CRDSA) [17] où un timeslot peut être utilisé par tous les terminaux ayant des données à transmettre, sans qu'il soit alloué préalablement à un Terminal en particulier.

Le faible volume de données permettait alors d'utiliser des méthodes d'accès à contention simples, de type ALOHA, dont le comportement à faible charge – peu de Terminaux s'enregistrent en même temps – est satisfaisant [18]. Leur utilisation pour la transmission de données a été introduite avec DVB-RCS2, en complément du DAMA, afin de diminuer les délais pour les applications interactives. Bien qu'efficaces, les volumes de données concernés par ces applications sont faibles, comparativement aux méthodes déterministes. En effet, ces méthodes ont un comportement moins bon à forte charge, à cause de la présence de nombreuses transmissions. Nous avons donc choisi, dans cette étude, de nous concentrer sur la partie DAMA de l'allocation de ressources.

Une des caractéristiques importantes du DAMA est qu'il nécessite une requête de la part du Terminal, en termes de volume ou de débit, informant la Passerelle de ses besoins. Dans DVB-RCS2, cette requête peut être exprimée selon deux catégories, selon les besoins du Terminal. On trouve également une allocation dite, *non sollicitée*, qui est allouée à chaque Terminal sans que celui-ci soit amené à demander ces ressources.

- Constant Rate Assignment (CRA) : allocation non sollicitée, utilisée pour les communications de type Voice over IP (VOIP) ou vidéoconférence. Cette allocation de ressources peut se faire par l'intermédiaire d'algorithmes de contrôle d'admission ou Call Admission Control (CAC), un processus plus lent que l'allocation de ressources. De ce fait, leur gestion est relativement aisée, puisqu'elle repose sur un simple calcul de capacité, et que l'allocation résultante est périodique. La dynamique de cette allocation est donc potentiellement plus lente que celle du processus de requêtes.
- Rate-Based Dynamic Capacity (RBDC) : demande en débit, dynamique, et pour une certaine période prédéterminée. Cette requête est utilisée pour écouler le trafic n'ayant pas un comportement facilement prédictible. Elle peut bénéficier de

l'agrégation de plusieurs flux, qui aura tendance à lisser le trafic.

- Volume-Based Dynamic Capacity (VBDC) : demande en volume, généralement utilisée pour un trafic aperiodique et peu volumineux, bien adaptée au trafic engendré par HyperText Transfer Protocol (HTTP).

Si le calcul du CRA et des requêtes VBDC est relativement aisé, il n'en est pas de même pour les requêtes RBDC, où le Terminal doit estimer le débit dont il aura besoin pendant la période de validité de sa requête. Cette période est bornée par le délai de propagation très important dû au satellite. De ce fait, les mises à jour de l'allocation en fonction de la requête ne pourront se faire qu'avec un retard d'au moins deux fois le délai de propagation, soit 500 ms. De plus, la mise à jour des requêtes s'effectue au moment du SYNC SLOT, timeslot réservé à la signalisation Terminal-Passerelle. Du fait des ressources limitées allouées à la signalisation, la période de ce SYNC SLOT, pour un Terminal, peut être bien plus grande que 500ms, et des valeurs d'une seconde sont courantes. Ce retard entre requête et allocation a un impact important, puisqu'il suppose d'estimer correctement la requête de capacité en RBDC, sur une période relativement longue par rapport au comportement des applications, tâche notoirement difficile [19].

1.3 Demande et Capacité

Les mécanismes d'ordonnancement dans les réseaux sans fil sont à la rencontre de deux processus, l'un provenant des couches hautes du réseau, que nous appellerons demande, et l'autre provenant des couches basses, que nous appellerons capacité. La demande émane des utilisateurs du système, elle se traduit par une charge entrante, ou un débit entrant, un volume de données à traiter en un certain temps tout en respectant certaines contraintes de QoS. La capacité représente la quantité de données que le système peut traiter par unité de temps, étant liée aux conditions de transmission. En raison du caractère aléatoire de ces dernières, et surtout du temps de propagation propre au système satellite, prédire la capacité s'avère bien plus complexe que dans les réseaux filaires. Toute la difficulté de l'ordonnancement vient du caractère fondamentalement décorrélé de ces deux processus, dont il faut tenir compte pour à la fois respecter la demande, et tirer le meilleur parti de la capacité.

1.3.1 Demande : Services et Utilisations

Le système ainsi décrit précédemment présente, du point de vue fonctionnel, une similitude flagrante avec les réseaux de type ADSL, où le réseau d'accès cuivre serait remplacé par un réseau d'accès Satellite. Les services offerts seront donc très similaires

aux offres de type *triple-play* que l'on peut trouver chez bon nombre d'opérateurs. Étudier la demande du système revient alors à étudier le comportement de ces services.

De tels systèmes existent actuellement, tels ceux utilisant le satellite Ka-Sat [20] d'Eutelsat, bien que les fonctionnalités offertes ainsi que la capacité n'égalent pas, pour un coût équivalent, celles des réseaux terrestres modernes.

On retiendra trois types de services, très similaires à ceux offerts dans les réseaux ADSL :

- L'accès à Internet, ou trafic Best Effort (BE). Il se caractérise par un comportement apériodique, ou en rafales (*bursty*), très différent du trafic vidéo. Son volume étant prépondérant sur la voie Aller, il aura une influence importante sur la performance de l'ordonnancement. Ce trafic inclut aussi bien la navigation web, que le transfert de fichiers, les activités interactives comme les jeux, la messagerie, et le visionnage de vidéos en ligne. Bien que ces applications aient un comportement différent, aucune différenciation n'est faite, et le trafic est considéré comme de même priorité. Aucune garantie n'est fournie sur ce trafic, puisque destiné à un réseau typiquement BE, Internet. Les paquets issus de ce type de trafic sont de taille variable [21].
- La voix sur IP (VOIP) ou la vidéoconférence. Sur la voie Aller, même si le volume de ce trafic est faible en comparaison du trafic BE, les contraintes en termes de temps d'attente et de gigue de ce trafic le rendent complexe à traiter. En effet, le long délai de propagation introduit par la liaison satellite dégrade la qualité de la communication ; il est donc nécessaire de le traiter en priorité de manière à ne pas le retarder au delà d'un seuil imposé. Bien qu'il possède un débit constant par conversation, les conversations VOIP ne sont pas permanentes, et il sera par conséquent impossible d'utiliser un ordonnancement périodique. Sur la voie Retour, nous verrons dans la suite que le trafic VOIP et vidéoconférence peut prendre une part significative du trafic demandé par un Terminal. De manière générale, on peut considérer que les paquets issus de ce trafic, pour une communication donnée, sont de taille constante.
- Le trafic vidéo. Bien que des services de diffusion de télévision puissent coexister avec notre système, il constituent deux éléments séparés et indépendants de notre système d'étude. Le trafic vidéo décrit ici pourrait correspondre à de la diffusion en continu (ou *streaming*) ou bien à un service de Video on Demand (VOD), tous deux proposés par l'opérateur, et donc clairement identifiés.

L'intérêt de cette architecture est d'appliquer une QOS adaptée à ce trafic, plutôt que de le considérer comme faisant partie du trafic BE. Les contraintes portent ici sur la gigue, qui peut fortement dégrader les performances, ainsi que sur le

débit très important demandé par certaines vidéos en haute définition. Cependant, ce trafic est relativement moins abondant que le trafic BE sur la voie Aller, et quasiment absent sur la voie retour. De même que pour la précédente catégorie, on observe peu de dispersion dans la taille des paquets engendrés par ce type de trafic.

La demande peut alors s'exprimer, dans le cas de la voie Aller, par la somme des débits requis par chaque application, auquel il faut ajouter les contraintes de QoS propres à chaque type de trafic. Outre le fait que cette demande est constituée de trafics de nature très différente, elle présente un caractère dynamique, dont les caractéristiques temporelles sont difficiles à exprimer. En effet, si la dynamique des appels téléphoniques est connue, de l'ordre de quelques minutes par appel et avec un débit constant, elle est très différente de celle du trafic HTTP, où l'ordre de grandeur est plutôt de quelques *bursts* de paquets par seconde, entrecoupés de longues périodes de silence.

En ce qui concerne la voie Retour, la demande subit les mêmes évolutions que la voie Aller, avec cependant la différence notable que cette demande passe par des requêtes de trafic (RBDC ou VBDC). Il y aura donc potentiellement un décalage entre la demande effective des Terminaux et les requêtes de trafic, issues d'estimations. On observera d'ailleurs un comportement de système bouclé, similaire à la boucle ACM, mais dont la variabilité, inhérente au trafic, sera supérieure. De plus, le débit relativement faible engendré par un Terminal, en comparaison du débit sur la voie Aller, rend très difficile sa prédiction. La demande aura donc tendance à évoluer plus rapidement sur la voie Retour, pour un Terminal donné. Et si la demande peut s'exprimer facilement à l'aide des requêtes de trafic, elle ne correspond qu'à une estimation de la demande réelle, et dépend fortement de la manière dont l'estimation est faite.

En résumé, la demande du système se caractérise par une évolution sur des échelles temporelles très différentes, et une nature de trafic disparate, des flux très réguliers, type VOIP, avec de fortes contraintes de QoS, aux flux irréguliers et BE, de type HTTP. Pourtant, étant donné le nombre important d'utilisateurs de ce service, on peut faire l'approximation que, d'un point de vue statistique, la charge moyenne globale évolue assez lentement pour qu'elle change peu à l'échelle de temps de l'ordonnancement ou de l'allocation de ressources. On peut alors donner une répartition statistique du trafic pour chaque classe, comme indiqué, pour la voie Aller, à la table 1.2. Cette charge correspond au trafic agrégé à destination des Terminaux.

Cette approximation est appuyée par le fait que la voie Aller comporte un débit important en comparaison de la voie Retour, rendant son comportement moyen plus facile à prédire, en observant le système sur une période longue devant les changements

Type de trafic	Charge
VoIP et Vidéoconf.	0.1
Vidéo	0.3
BE	0.6

TABLE 1.2 Exemple de répartition du trafic pour la voie Aller

météorologiques. Il sera en revanche impossible de prédire le comportement exact de chaque flux, à chaque instant, comme c'est le cas dans les systèmes de diffusion de télévision.

1.3.2 Capacité

La capacité d'un Système de Télécommunications par Satellite peut s'exprimer selon de nombreuses métriques. D'un point de vue physique, cette capacité s'exprime en termes de bande passante allouée. Pour une porteuse classique, on donne généralement une bande passante, typiquement $B = 36\text{MHz}$ [22]. Mais la relation entre ce chiffre et le débit binaire effectif atteignable sur cette porteuse n'est pas immédiate. Dans le cas d'un système sans ACM, où modulation et codage sont fixés, le débit est donné par :

$$R_b = \rho \log_2(M) \frac{B}{1 + \alpha} \quad (1.1)$$

Où

- R_b est le débit binaire (bits/s)
- α est le facteur d'amortissement du filtre d'émission ($\alpha \in [0, 1]$)
- ρ est le taux de codage
- M est l'ordre de la Modulation
- B est la bande passante du système (Hz)

Afin de garantir un TEB de référence, ρ et M sont calculés de manière à fournir ce niveau de service avec une disponibilité temporelle fixée (par exemple 99%). Cette technique *pire cas* ne permet pas d'utiliser de manière optimale la bande passante allouée au système.

Les systèmes avec ACM permettent d'adapter ρ et M , en d'autres termes le MODCOD, aux conditions de transmission rapportées par les Terminaux. En effet, les fréquences utilisées dans les Systèmes de Télécommunication par Satellite sont très sensibles à l'atténuation de l'eau sous forme gazeuse ou liquide présente dans l'atmosphère. Les Terminaux, ayant une puissance limitée et une antenne de faible diamètre, subiront donc

une atténuation importante en cas de pluie. Cette atténuation entraînant des erreurs de transmission, il est donc nécessaire d'adapter la protection des données contre ces erreurs, dans l'objectif de fournir un TEB constant. Ces fréquences, situées dans la bande Ka, se situent entre 22 et 40 GHz.

La figure 1.7 présente, de manière qualitative, l'évolution du débit (et donc du MODCOD) en fonction des conditions météorologiques qui affectent le Terminal.

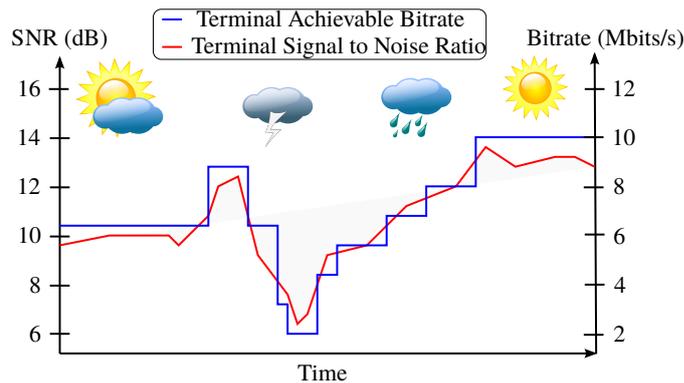


FIGURE 1.7 Relation entre conditions météo et débit pour un Terminal

La durée moyenne d'un événement de pluie est de quelques minutes [23], ainsi la variation du MODCOD d'un Terminal aura-t-elle une durée caractéristique du même ordre de grandeur. C'est une différence fondamentale des Systèmes de Télécommunication par Satellite par comparaison aux systèmes terrestres, où les évanouissements (*fading*) ont pour ordre de grandeur quelques centièmes de seconde, voire moins. Ces évanouissements sont dus au canal de propagation mobile, et au multi-trajet.

Si la bande passante est connue, il est néanmoins complexe de définir la capacité du système comme débit total, puisqu'elle dépend des Terminaux considérés, et évoluera ainsi au rythme des changements de MODCOD. Cependant, étant donné le nombre très important de Terminaux considérés (quelques milliers) il est possible de faire une approximation statistique similaire à la demande. De fait, même si les conditions météorologiques au-dessus d'un Terminal donné évoluent, celles sous l'ensemble de la couverture sont relativement stables, puisqu'elles considèrent une zone géographique importante, tout du moins à une échelle temporelle grande par rapport à l'allocation de ressources ou l'ordonnancement.

Dès lors, on peut considérer que, si les Terminaux changent de MODCOD, la répartition statistique de ces derniers reste relativement stable sous l'ensemble de la couverture du Satellite, pendant un temps significatif pour notre étude, généralement quelques dizaines

de minutes. Une répartition de type *ciel clair*, où la majorité des Terminaux ont un MODCOD relativement fort, est exposée dans la table 1.3 [24].

#	MODCOD	Charge
0	QPSK 2/3	0.005
1	QPSK 3/4	0.005
2	QPSK 4/5	0.005
3	QPSK 5/6	0.01
4	8-PSK 3/5	0.28
5	8-PSK 2/3	0.58
6	8-PSK 2/3	0.10
7	16-APSK 2/3	0.01
8	16-APSK 3/4	0.005

TABLE 1.3 Exemple de répartition des MODCOD de type ciel clair

On voit ici apparaître un problème classique de l'ordonnancement : si trop d'importance est accordée aux utilisateurs ayant un faible MODCOD, la capacité du système s'en trouve réduite (le débit total de sortie est faible, cf table 1.1). Si elle est accordée aux utilisateurs ayant un MODCOD dans les plus favorables, la capacité est bien plus importante, mais on crée un risque de famine des utilisateurs ayant un faible MODCOD. La tâche de l'ordonnanceur s'en trouve donc plus complexe, puisqu'il faut trouver un compromis entre ces différents MODCOD et l'efficacité de l'utilisation de la bande passante, point crucial au regard du coût de l'utilisation des fréquences satellite.

De plus, si la répartition des MODCOD est stable à court terme, elle peut changer significativement selon le type d'évènement météorologique [23], il faudra alors que l'algorithme soit capable de s'adapter à cette nouvelle distribution, et d'offrir des performances comparables dans des scénarios très différents.

1.4 Conclusions de l'étude

Dans ce chapitre, nous avons présenté le contexte de notre étude des Systèmes de Télécommunication par Satellite, d'un point de vue fonctionnel et protocolaire. Nous avons choisi de porter notre attention sur la couche accès de ces systèmes, du fait des profonds changements introduits par les dernières générations des standards de télécommunications satellitaires, qui obligent à revoir leur mise en œuvre en profondeur. Nous avons également décrit les deux processus majeurs interagissant avec la couche accès, nommés demande et capacité. Suite à cette présentation, nous pouvons en tirer les enseignements présentés

ci-dessous.

Premièrement, les caractéristiques de la voie Aller et de la voie Retour, telles que définies ici, apparaissent comme très différentes. En effet, là où la voie Aller est confrontée à un problème d'ordonnancement classique, bien que rendu complexe par l'encapsulation GSE dans les BBFRAME, la voie retour constitue un problème typique d'allocation de ressources, qui se place à une échelle plus lente que celle de l'ordonnancement. De ce fait, il semble approprié, au premier abord, d'en séparer le traitement, afin de pouvoir proposer un algorithme adapté à chaque voie. Nous aborderons donc, dans cette thèse, voie Aller et voie Retour de manière successive.

De plus, la forte variabilité du trafic est un facteur important dans les Systèmes de Télécommunications par Satellite : plutôt qu'un trafic uniforme et régulier, nous avons à présent affaire à un trafic issu de multiples sources, aux caractéristiques disparates (débit, QoS, taille de paquets, etc). Ce changement nous oblige à repenser les architectures des systèmes d'ordonnancement et d'allocation de ressources, afin de pouvoir prendre en compte cette variabilité. Ceci est manifeste pour le passage de MPE à GSE, qui non content de prendre en compte de nombreux flux de QoS très différentes, permet une encapsulation aisée d'IP et des paquets de taille variable de manière générale, en minimisant l'overhead. Dans cette thèse, nous accorderons donc une importance particulière aux mécanismes de QoS, leur mise en place ainsi que leur application. Ces garanties sont absolument nécessaires au regard de la disparité du trafic, ainsi que des contraintes très strictes du trafic VOIP et vidéoconférence, dans le cas de la voie Aller comme dans celui de la voie Retour. Nous observerons également comment traiter efficacement la variabilité de la taille des paquets.

Enfin, dans l'optique de proposer une solution crédible pour une implantation réelle, nous serons notamment attentifs aux problèmes de complexité algorithmique, d'implantation et de performance temps réel.

ÉTAT DE L'ART DES SOLUTIONS D'ORDONNANCEMENT ET D'ALLOCATION DE RESSOURCES

L'ordonnement aussi bien que l'allocation de ressources ont suscité de nombreuses recherches, du fait de leur importance dans l'amélioration des performances des systèmes. Les réseaux sans fil, et plus particulièrement les Systèmes de Télécommunication par Satellite n'ont pas échappé à cet engouement, même s'ils présentent des caractéristiques bien distinctes, et constituent donc un domaine d'application à part entière.

Comme décrit dans le chapitre précédent, si l'ordonnement de flux homogènes (MPEG-2 dans le cas du satellite) est aisé, l'introduction de flux aux caractéristiques très différentes, ainsi que versatilité temporelle du support radio le complexifient grandement. Ce problème n'est pas inhérent au satellite, mais les caractéristiques du système sont très contraignantes, notamment quant à l'évolution conjointe de la capacité et de la demande, et aux méthodes d'encapsulation, et doivent être prises en compte dans la conception des algorithmes.

Notre objectif dans ce chapitre est de présenter, de manière synthétique, les principaux aspects des solutions d'ordonnement et d'allocation de ressources qui ont été développées, pour les systèmes satellites en particulier. Nous aborderons les solutions proposées dans la littérature à l'aune du système défini dans le précédent Chapitre, en prenant les architectures DVB-S2 et DVB-RCS2 comme références. Une partie importante de notre étude sera donc consacrée à déterminer comment adapter des solutions présentées à notre système de référence. Notons que, bien que très différentes, les solutions conçues pour les systèmes terrestres pourront également être étudiées, si leur adaptation semble

pertinente. Du fait du caractère éminemment dynamique et décorrélé de la capacité et de la demande du système étudié, nous porterons une attention particulière à la façon dont les solutions proposées dans la littérature prennent en compte ces paramètres.

La plupart des algorithmes d'ordonnement utilisent une ou plusieurs métriques afin d'évaluer l'état du système et prendre des décisions en conséquence. Nous porterons donc tout d'abord succinctement notre attention sur l'estimation de ces métriques. Nous présentons dans un deuxième temps une classification générale des méthodes d'ordonnement où trois outils mathématiques, largement plébiscités dans la littérature, sont exposés. Puis, un état de l'art des méthodes développées pour le satellite est présenté. Cette étude permet de mettre en évidence les difficultés liées à notre problème, ainsi que les insuffisances des solutions proposées jusque-là.

2.1 Conventions de nommage et métriques

L'utilisation de métriques pour observer l'état d'un système de télécommunication est nécessaire dès lors que ce système évolue dans le temps, qui plus est de manière difficilement prédictible. Une métrique est une valeur numérique permettant de représenter l'état du système. Cette définition comporte également un aspect temporel, on retrouvera des métriques moyennes, définies sur une certaine période de temps, ou *a posteriori*, et des métriques instantanées, dont la valeur n'a de sens qu'à un instant donné. On peut citer par exemple, le débit moyen observé par un Terminal ou délai d'une trame dans une file d'attente de la Passerelle. Cependant, l'estimation de ces métriques pose problème : elle doit prendre le moins de ressources (mémoire, calcul) possibles, tout en étant pertinente. Nous proposons donc ici un compromis entre ces deux objectifs.

2.1.1 Définitions

Nous définissons tout d'abord l'*instant d'ordonnement*, comme le moment où la dernière BBFRAME est transmise pour la voie Aller. Cet instant revient de manière régulière, plus précisément périodique pour la voie Retour, dépendant du temps de transmission de la BBFRAME pour la voie Aller. Disposant d'une numérotation incrémentale n des BBFRAME selon une origine déterminée, nous notons $t(n)$ l'instant auquel la BBFRAME n commence sa transmission. Nous définissons ainsi une mesure discrète du temps, $t(0), t(1), \dots, t(n)$, qui servira d'horloge à notre système. Pour plus de clarté, on utilisera de manière interchangeable l'instant $t(n)$ et n dans la suite de cette discussion. Cette modélisation est présentée à la figure 2.1.

Il semble important de noter que cette représentation est non seulement arbitraire,

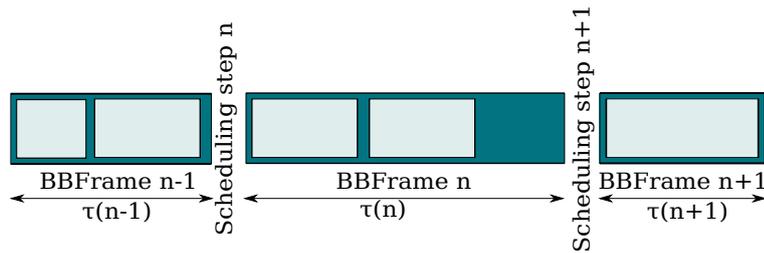


FIGURE 2.1 Mesure discrète du temps

mais dépend aussi fortement du système considéré (fréquence des BBFRAME, MODCOD), bien qu'adoptée dans de nombreux travaux [25] [26] [27] [28] [29] [5]. On dit également que l'algorithme est auto-temporisé (*self-clocked*, [30]) car il n'utilise pas d'horloge externe par opposition à des algorithmes tels que Weighted Fair Queuing (WFQ)[31] [32]. Sur la voie Retour, cette mesure du temps correspond à la période des Supertrames DVB-RCS2.

S'appuyant sur ces valeurs, plus précisément sur l'état du système observé à l'instant n (l'ensemble des files d'attente et des variables associées de la couche accès), plusieurs *solutions d'ordonnancement* sont déterminées. Elles correspondent aux différentes possibilités dont dispose la couche accès pour ordonnancer ou allouer les ressources. On appellera *décision d'ordonnancement* la sélection de la solution d'ordonnancement (respectivement d'allocation de ressource pour la voie Retour). La façon de prendre cette décision dépend bien évidemment de l'algorithme, ou une *évaluation* sera faite de chaque solution d'ordonnancement, selon des critères propres à l'algorithme. L'outil permettant de faire cette décision est appelé *critère d'ordonnancement*.

Enfin, on définit par *scénario d'ordonnancement* l'ensemble des valeurs des paramètres d'influence pour l'ordonnancement dans notre système, à savoir :

- le nombre de Terminaux ;
- le type de trafic, ses caractéristiques, et la voie considérée (Aller ou Retour) ;
- la répartition des MODCOD des Terminaux.

En résumé, tout changement dans les paramètres de la demande ou de la capacité telles que définies au chapitre précédent constitue, pour notre système, un scénario différent.

2.1.2 Métriques

Si l'algorithme considéré fait usage de valeurs moyennes, la valeur de la variable moyenne \bar{x} utilisée à l'instant n sera $\bar{x}(n)$, et la valeur instantanée mesurée sur le système sera notée $x(n)$. On dira que l'algorithme *valorise* la métrique x si la valeur de celle-ci est utilisée dans la décision d'ordonnancement, voire dans l'expression du critère

d'ordonnancement.

Les instants d'observation du système correspondant à ceux d'ordonnancement, l'algorithme n'a donc accès qu'aux valeurs instantanées mesurées, $x(n)$. Il existe plusieurs manières d'estimer la valeur moyenne de x à partir de ces valeurs instantanées, nous avons choisi la moyenne exponentielle, qui permet de filtrer les événements ayant une forte amplitude par rapport à la moyenne, lissant ainsi la mesure. Ce type de mesure, très largement employé, offre un compromis entre précision et simplicité de calcul. En considérant la métrique \bar{x} , sa moyenne exponentielle est donnée par :

$$\bar{x}(n+1) = \alpha\bar{x}(n) + (1-\alpha)x(n) \quad (2.1)$$

Où α est un facteur d'amortissement, proche de 1. On remarque que cette mesure ne dépend pas du temps écoulé entre deux instants, ce qui peut être préjudiciable quand cette durée est variable, par exemple sur la voie Aller, lorsque deux BBFRAME de MODCOD différent se suivent. Elles n'auront pas le même temps de transmission, le temps entre deux instants sera donc variable. Les auteurs de [28] proposent d'adapter α à la durée entre deux instants, en posant :

$$\alpha(n) = \alpha^{\tau(n)} \quad (2.2)$$

Où $\tau(n)$ est la durée entre deux instants d'ordonnancement, à savoir le temps d'émission de la BBFRAME choisie. Cette formule sera utilisée par défaut dans la suite de notre discussion.

2.2 Outils analytiques d'ordonnancement

Dans cette partie, nous exposons trois méthodes permettant d'évaluer les solutions d'ordonnancement ou d'allocation de ressources, chacune prenant appui sur une modélisation mathématique différente. Notre objectif, au-delà de présenter une classification, est de montrer comment des méthodes a priori très différentes peuvent mener au même résultat, à travers un exemple commun. Notre système de référence est constitué par la couche accès de la Passerelle, sur la voie Aller ou bien sur la voie Retour. L'ordonnancement est l'algorithme de référence de cette étude, mais les explications données ici sont, par défaut, valables également pour la voie Aller. Les deux processus sont distingués lorsque ce n'est plus le cas.

Nous considérons ici un nombre quelconque d'utilisateurs, indexés par la variable i . Par souci de généralité, nous n'explicitons pas davantage ces utilisateurs, qui peuvent être aussi bien des files d'attente situées dans la Passerelle, que des Terminaux ou bien

encore des flux. Nous considérons une métrique générique, notée $x(n)$, valeur moyenne ou instantanée, représentant l'état du système, et nous supposons de plus qu'il est possible de définir clairement, et pour chacun de ces utilisateurs, un état $x_i(n)$. On peut citer comme exemple, le débit moyen, ou bien le taux de perte. Cette valeur est disponible instantanément pour le système de référence, et mise à jour à chaque nouvel instant d'ordonnement.

2.2.1 Critère Explicite

Cette catégorie de solutions regroupe un très grand nombre d'algorithmes développés pour l'ordonnement dans les systèmes sans fil. Leur principe peut être résumé de la façon suivante :

- Sélectionner une ou plusieurs métriques, x_i représentatives de l'état de chaque utilisateur dans le système.
- Définir une évaluation de cet état à chaque instant, $f_i(x_i(n))$, que l'on associera au critère d'ordonnement.
- Évaluer les différentes solutions disponibles d'ordonnement ou d'allocation de ressources, et prendre la décision d'ordonnement.

Définition empirique

Cette définition reste générale, et a pu donner lieu à des algorithmes très différents, [5],[33],[34],[4]. Elle est dite explicite, ou empirique, car l'évaluation utilise une fonction f qui doit être explicitement donnée, et qui repose généralement sur des considérations empiriques. De même, la manière d'utiliser le critère ainsi défini doit être également explicitée.

Dans le cas où un seul utilisateur est servi à chaque instant d'ordonnement, un exemple très simple d'algorithme d'ordonnement construit sur ce critère est donné par :

$$i^* = \arg \max_i f_i(x_i(n))$$

Autrement dit, l'utilisateur i^* est servi à l'instant n seulement si la valeur de son critère est la plus élevée. Cette formulation a l'avantage d'être simple et facile à implanter, un simple calcul suffit à déterminer le prochain utilisateur à servir. De plus, sa complexité croît linéairement avec le nombre d'utilisateurs, et son implantation peut être facilement parallélisée.

Une autre utilisation possible de ce critère peut se matérialiser au travers d'un algorithme d'ordonnement classique, tel que le RR, où chaque utilisateur est servi

successivement également, selon un ordre défini. Ici, on peut supposer les utilisateurs étant triés par critère décroissant, cet ordre étant mis à jour régulièrement, avec les nouvelles valeurs du critère. Dans un système ne comportant que des paquets de taille constante, le RR se trouve être équitable et performant[35], et présente une complexité algorithmique constante, $o(1)$.

On peut aussi envisager d'utiliser ce critère dans les variantes à priorité de RR. Le Weighted Round Robin (WRR) permet de différencier le traitement entre les différents utilisateurs, en leur assignant un poids. Ce poids peut correspondre, par exemple, à une contrainte de QOS. Principalement développé pour ATM, [36] [37], le poids peut être exprimé comme le temps relatif accordé à un utilisateur lors d'un cycle du RR. Cet algorithme présente un intérêt certain lorsqu'il est utilisé avec des paquets de taille constante, comme c'est le cas des cellules ATM, puisque son comportement peut être prédit[4].

DRR, présenté dans [38], est une variante du WRR adaptée aux paquets de taille variable. Le déficit représente ainsi le volume alloué à chaque utilisateur à chaque tour du RR. Ce déficit peut être également adapté en fonction de contraintes de QOS. Ce déficit est consommé lorsque l'utilisateur a des paquets à transmettre, et cumulé lorsque le déficit accumulé n'est pas suffisant pour transmettre le prochain paquet. De cette manière, le DRR permet également un lissage du trafic. Tout comme RR et WRR, la complexité algorithmique de DRR est faible ($o(1)$).

Nous pouvons enfin citer le WFQ, [39], bien que cet algorithme soit plus complexe. Il repose sur un partage équitable de la capacité entre les utilisateurs, selon un modèle de partage du temps d'exécution d'un processeur entre plusieurs tâches. Cet algorithme a suscité nombre de recherches [40] [41] [42] [43] [44]. Le fait que cet algorithme suppose une horloge externe au système, ainsi que sa complexité, le rendent difficile à utiliser pour un système dont la capacité varie avec le temps. De plus, il possède une complexité algorithmique supérieure, $o(\log(N))$.

En résumé, cette formulation permet une très grande flexibilité, ainsi qu'une faible complexité d'implantation, deux atouts très importants pour des systèmes comme une Passerelle satellite, où les décisions d'ordonnancement doivent être effectuées en quelques microsecondes. Mais cette formulation très simple cache cependant une grande complexité de conception, dans le choix de f mais aussi dans l'utilisation qui est faite du critère.

En effet, puisqu'il n'existe pas de contrainte particulière sur ces deux choix, le nombre des solutions possibles est gigantesque. Pourtant, le travail d'évaluation de chaque critère est complexe : une évaluation analytique est possible pour un critère simple [45], mais devient trop complexe si l'expression du critère contient plusieurs métriques différentes,

chacune ayant son propre impact sur f . Il est alors difficile d'étudier l'influence à la fois des métriques et du paramétrage, et donc de déterminer la solution optimale. L'évaluation à l'aide de simulations souffre des mêmes difficultés, d'autant plus que les performances se révèlent très différentes selon le scénario ou le paramétrage.

Règles d'ordonnement

Les règles d'ordonnement sont une réponse partielle aux problèmes des critères explicites cités précédemment. On appelle règle d'ordonnement une catégorie particulière de critères explicites.

Ces règles prennent la forme de critères dont l'expression, tirée de travaux de référence [46] [27] [47], est définie, permettant ainsi une formulation plus homogène des algorithmes d'ordonnement [48].

Les paramètres utilisés par ces règles sont des valeurs moyennes de métriques classiques, telles que le débit sortant, le délai du paquet en tête de file d'attente (Head of Line (HOL)), ou encore la qualité instantanée du support.

Une règle particulièrement populaire est Proportional Fairness (PF), dont l'expression est la suivante :

$$i^* = \arg \max_i \frac{r_i(n)}{\bar{r}_i(n)} \quad (2.3)$$

Où :

- r_i est le débit instantané atteignable par l'utilisateur i , dépendant des conditions de transmission ;
- \bar{r}_i est le débit moyen obtenu par l'utilisateur i .

PF réalise un compromis entre efficacité spectrale et équité. Par l'intermédiaire de la prise en compte de la qualité instantanée du canal, PF privilégie les utilisateurs ayant un bon débit atteignable, ce que l'on nomme un comportement *opportuniste*. Une règle ne prenant en compte que ce seul paramètre permettrait de maximiser l'efficacité spectrale, pourvu que les utilisateurs servis aient toujours des données à transmettre ou à recevoir. Il dégraderait cependant l'équité puisqu'un utilisateur ayant de mauvaises conditions de transmission pendant un temps assez long pourrait ne pas être servi pendant tout ce temps, créant un phénomène de famine. C'est pourquoi cette influence est contrebalancée par la prise en compte du débit moyen obtenu. Ainsi, même si un utilisateur possède de très bonnes conditions de transmission, la valeur de son critère chutera au fur et à mesure qu'il sera servi, le rendant ainsi moins prioritaire.

Cette règle s'adapte facilement à la voie Aller aussi bien qu'à la voie Retour : il est possible de choisir un ou plusieurs utilisateurs, triés par critère PF décroissant. Elle est

majoritairement adaptée au trafic dit élastique, c'est-à-dire dont les besoins en bande passante peuvent s'adapter à la capacité disponible. Le trafic inélastique, généralement à débit constant (VOIP, vidéo), nécessite une règle plus complexe, comme par exemple Modified Latest Weighted Deadline First (M-LWDF)[27] :

$$i^* = \arg \max_i \gamma_i \bar{w}_i(n) r_i(n) \quad (2.4)$$

Où

- γ_i est une priorité de QoS ;
- $\bar{w}_i(n)$ est le temps d'attente moyen en tête de file de l'utilisateur i .

Cette règle permet un compromis entre débit, au travers de r_i , et le délai. On peut également citer la règle exponentielle, ou Exponential Rule - PF (EXP-PF) [47], ou encore Earliest Deadline First (EDF) [49], tous deux adaptés au trafic temps réel.

Les règles d'ordonnancement semblent fournir un bon compromis entre complexité, performance et facilité de conception. L'abondance des travaux de recherche sur ces critères spécifiques est également un avantage certain.

On remarque cependant deux inconvénients à l'utilisation des règles d'ordonnancement. Premièrement, il est difficile d'obtenir des résultats de stabilité de ces règles, autrement que par des simulations étendues, prenant en compte de très nombreux cas. Cela peut se révéler critique dans des systèmes opérant à forte charge, où le risque de déstabilisation du système est grand. Deuxièmement, les algorithmes d'ordonnancement définis ici sont dits *myopes*, puisqu'ils ne considèrent, pour prendre leur décision, que le passé du système. Autrement dit, il n'est pas possible de prédire facilement l'impact des décisions d'ordonnancement sur le système, et ainsi de savoir l'état vers lequel l'ordonnancement mène le système. Cela peut être particulièrement préjudiciable pour un système évoluant avec le temps : le critère peut être stable sous un certain scénario mais devenir instable dès que ce scénario change significativement. Cette propriété marque aussi un frein au développement des règles d'ordonnancement, et particulièrement dans le cas de règles ajustées pour obtenir de bonnes performances dans un scénario particulier, puisqu'il n'y a aucune garantie que cet ajustement soit aussi efficace avec un autre scénario.

2.2.2 Théorie des jeux

Afin de donner un cadre mathématique rigoureux à l'ordonnancement, de nombreux travaux ont eu recours à la Théorie des Jeux pour modéliser des situations où plusieurs utilisateurs se partagent un même système de télécommunications. La Théorie des Jeux a été introduite au milieu du XXe siècle, afin de formaliser les situations de compétition

entre acteurs économiques. Ces acteurs, appelés Joueurs, prennent des décisions selon une *stratégie* de manière à maximiser leurs *gains*. Appliquée aux réseaux sans fil, on peut aisément voir comment il est possible d'appliquer cette théorie, par exemple pour modéliser les protocoles d'accès à contention, où le gain serait le débit reçu en fonction de la stratégie, qui serait elle la manière avec laquelle l'utilisateur accède au support.

Équilibre de Nash

Un des résultats majeurs de la Théorie des Jeux est l'Équilibre de Nash, introduit par John Nash [50]. Il stipule que, dans le cas de joueurs ayant une estimation rationnelle de leurs gains en fonction de leur stratégie, et ayant une connaissance complète des stratégies des autres joueurs, chaque joueur atteindra un équilibre où son gain sera maximal, sachant les stratégies adoptées par les autres joueurs. En d'autres termes, dans cet équilibre, chaque joueur prend en compte les stratégies de tous les autres joueurs, et décide de la meilleure stratégie à adopter. Cet état est un équilibre, puisque aucun joueur ne peut augmenter son gain en changeant de stratégie [51]. On appelle cette solution Nash Bargaining Solution (NBS).

On peut formuler un NBS de la manière suivante, pour des utilisateurs dont le gain est le débit moyen reçu \bar{r}_i , l'Équilibre de Nash est solution du problème d'optimisation suivant :

$$\begin{aligned} \max \prod_i \bar{r}_i \\ \text{avec } \sum_i \bar{r}_i \leq C \end{aligned} \tag{2.5}$$

Où C est la capacité totale du système, exprimée ici en termes de débit. Cette formulation est particulièrement utile pour les problèmes de partage de capacité [52], comme dans les réseaux filaires, ou bien dans un contexte statique, par exemple pour le dimensionnement dans les réseaux Orthogonal Frequency Division Multiple Access (OFDMA) [53]. De plus, la Théorie des Jeux apporte des garanties quand à la stabilité de l'équilibre atteint, ce qui est un avantage certain par rapport aux critères explicites. Nous noterons par la suite \bar{r}^* la solution de ce problème.

Il reste cependant un problème de taille avant de pouvoir utiliser la théorie des jeux de manière efficace : adapter cette formulation au caractère dynamique de notre problème. En effet, la formulation précédente donne un équilibre sur des valeurs moyennes, indépendantes du temps. Or, non seulement l'ordonnement ne peut se traiter sans dimension temporelle, mais il est même nécessaire de l'adapter aux évolutions du système. Par exemple, nous avons exprimé ici la capacité en débit. Comme vu dans le Chapitre 1,

cette capacité évolue au cours du temps, il serait donc plus juste de remplacer C par $C(n)$, ce qui a pour conséquence de modifier la solution de (2.5), puisque la contrainte de capacité change. Du reste, même en explicitant les valeurs de \bar{r}^* , comment faire le lien entre cette valeur et la décision d'ordonnement prise à l'instant n ? Ou encore, comment se traduit la valorisation du débit comme gain dans l'ordonnement? La Théorie des Jeux n'apporte ici aucune réponse, du moins en adoptant la formulation de (2.5), et le rapport entre l'équilibre découlant du NBS et la stratégie effective qu'il faut adopter à chaque étape d'ordonnement n'est absolument pas explicite. L'avantage apporté par la théorie des jeux se trouve donc ici mitigé : bien qu'elle fournisse des outils mathématiques rigoureux, sa formulation est peu adaptée à la dynamique du système, en particulier à celle de l'ordonnement.

De plus, en supposant qu'une méthode itérative ait été trouvée afin d'atteindre l'équilibre défini, le caractère à la fois dynamique et limité des ressources sera ici aussi une contrainte. En effet, si on suppose à l'instant n un état du système, c'est-à-dire des valeurs moyennes de débit $\bar{r}(n)$, il ne sera probablement pas possible d'atteindre \bar{r}^* en une seule étape d'ordonnement, car la vitesse à laquelle le débit moyen peut évoluer est limitée par la capacité $C(n)$. En d'autres termes, l'ensemble des états atteignables en une étape d'ordonnement est limité, et cet ensemble ne contient pas nécessairement l'optimal. Par conséquent, il sera nécessaire d'effectuer plusieurs étapes d'ordonnement avant d'atteindre l'équilibre défini. Or, cet équilibre, comme vu plus haut, dépend du temps lui aussi, par l'intermédiaire de $C(n)$: on risque alors de ne jamais pouvoir atteindre l'équilibre défini, puisqu'il évolue dans le temps. Formellement, si l'algorithme proposé converge vers l'équilibre en T étapes, il faudra que $C(n) = C(n + T)$ pour que l'équilibre soit effectivement atteint, faute de quoi le système n'atteindra pas l'équilibre à $n + T$, puisque ce dernier aura changé.

Proportionnal Fairness

Une propriété intéressante de cette formulation est la possibilité de retrouver le critère PF comme précédemment présenté [54]. En effet, en considérant le problème suivant, après passage au logarithme, équivalent à (2.5) :

$$\begin{aligned} & \max \sum_i \ln(\bar{r}_i) \\ & \text{avec } \sum_i \bar{r}_i \leq C(n) \end{aligned} \tag{2.6}$$

On obtient alors un problème d'optimisation non linéaire, strictement convexe, qu'il

est donc possible de résoudre à l'aide de méthodes Lagrangiennes. La solution de ce problème [55], qui n'est autre que \bar{r}^* , est dite *proportionnellement équitable*, du fait de la propriété suivante, obtenue pour $n \rightarrow +\infty$:

$$\sum_i \frac{r_i - \bar{r}_i^*}{\bar{r}_i} \leq 0 \quad (2.7)$$

où r_i est une allocation possible des débits moyens, ce qui signifie que le gain global apporté par un changement, par rapport au NBS, dans le débit d'un des utilisateurs (par exemple son augmentation) est nul ou négatif. Cette propriété assure un comportement équitable, empêchant un seul utilisateur d'accaparer la majorité des ressources : les joueurs vont coopérer afin d'arriver au meilleur compromis. Dans son application à des problèmes d'ordonnement, un algorithme du gradient a été adopté [26],[56], pour résoudre le problème de manière itérative. L'expression du critère d'ordonnement découlant de ce problème est la suivante :

$$i^* = \arg \max_i \frac{r_i(n)}{\bar{r}_i(n)} \quad (2.8)$$

Soit exactement l'expression (2.3), ce qui signifie que ce critère est non seulement équitable, mais également optimal pour un trafic élastique, valorisant le débit [57]. Il est important de noter qu'ici, la décision d'ordonnement ne se fait pas de manière myope : on choisit celle qui mène le plus rapidement possible à l'équilibre, ce qui est une propriété importante du fait de la dynamique du système. Cette méthode permet également d'évaluer selon une métrique définie, à savoir le gain à chaque étape, les différentes solutions d'ordonnement possibles.

Conclusion sur l'utilisation de la théorie des jeux

La théorie des jeux permet de placer l'ordonnement et l'allocation de ressources dans un cadre mathématique rigoureux, apportant ainsi des garanties d'optimalité ainsi que de stabilité autour d'un équilibre. Contrairement aux critères explicites exposés précédemment, la théorie des jeux définit clairement un objectif, une direction, vers laquelle l'ordonnement va se diriger. On est alors en mesure de déterminer et d'étudier cet équilibre, et ainsi de mieux connaître le comportement du système. Mais cette définition d'un équilibre, et donc de valeurs moyennes, n'apporte pas directement la solution au problème d'ordonnement, ni même à celui d'allocation de ressources, du fait de la dynamique rapide du système. La solution classique à ce problème est de déplacer le système, à chaque étape d'ordonnement, vers la solution du problème

posé. De ce fait, les avantages inhérents à la théorie des jeux semblent s'effacer devant la complexité de mise en place de la solution, ce qui est d'autant plus évident avec l'exemple de PF.

2.2.3 Fonctions d'utilité

Les fonctions d'utilité sont liées à la théorie des jeux, mais peuvent constituer un outil à part entière, utilisable sans avoir à déployer la formalisation nécessaire à la théorie des jeux. Formellement, une fonction d'utilité associe une variable, ou une métrique, à une utilité. Elle décrit donc comment la valeur de la variable ou de la métrique est valorisée par l'algorithme. On peut alors définir, pour notre système de télécommunications de référence, un ensemble de variables et leur utilité associée : on pourrait par exemple mesurer la satisfaction des utilisateurs vis-à-vis d'un certain niveau de service reçu, et faire en sorte que l'ordonnement maximise cette valeur.

Au premier abord, cette formulation ne présente pas d'avantage concret sur la théorie des jeux, puisque aucune notion d'équilibre ni d'objectif n'est présente dans cette formulation. Elle ne présuppose pas de comportement particulier des utilisateurs, ni même d'objectif à long terme en soi. Pourtant, cette solution a été retenue dans nombre de méthodes et d'algorithmes d'ordonnement [24],[58],[29],[59],[55],[60]. Ce succès s'explique par la flexibilité offerte par les fonctions d'utilité comparativement à la théorie des jeux, et par le fait qu'il est souvent possible d'obtenir des résultats comparables, sans faire appel à une théorie complexe et parfois contraignante.

Proportionnal Fairness dérivé de fonctions d'utilité

Supposons à nouveau des utilisateurs ayant un trafic élastique, dont le débit moyen alloué par le système est \bar{r} , et l'utilité correspondante à ce trafic $u(\bar{r})$. Si l'utilité représente la satisfaction de l'utilisateur vis-à-vis du service reçu, celle-ci devra être strictement croissante avec le débit pour un trafic élastique. Cependant, afin de différencier les débits faibles des débits importants, on peut poser $u(\bar{r}) = \ln(\bar{r})$. De cette manière, la différence entre deux allocations tend à diminuer au fur et à mesure que ces allocations augmentent, ce qui signifie qu'il sera plus *utile* d'augmenter le débit des utilisateurs ayant un débit moyen faible, plutôt que celui de ceux disposant déjà d'un débit moyen important.

L'objectif du système est relativement simple, il consiste à maximiser l'utilité du

système, à savoir la somme des utilités des utilisateurs :

$$\begin{aligned} & \max \sum_i \ln(\bar{r}_i) \\ & \text{avec } \sum_i \bar{r}_i \leq C(n) \end{aligned} \tag{2.9}$$

Cette définition, directement sous forme d'un problème d'optimisation, n'a pas eu besoin d'autres hypothèses que la convexité du problème, aisément démontrable, et la concavité de u . Nous retrouvons donc une formulation strictement équivalente à (2.6), à laquelle nous appliquerons la même résolution, pour retrouver le critère PF, dont l'expression est donnée par (2.3).

Compromis nécessaires

Comme nous l'avons vu avec l'exemple de PF, les fonctions d'utilité nécessitent de fournir un objectif, ici la maximisation de l'utilité du système. Une fois de plus, ce problème est défini à l'aide de valeurs moyennes, et posera donc les mêmes difficultés que la théorie des jeux quant à la résolution et à l'algorithme découlant de cette résolution.

L'avantage des fonctions d'utilité réside dans la flexibilité apportée : l'objectif de l'ordonnancement est explicitement défini, et il est possible de concevoir une fonction d'utilité plus complexe, qui représente déjà un compromis entre plusieurs variables, par exemple le délai et la gigue, ou encore le taux de perte. Cette flexibilité accrue a cependant un coût, puisqu'elle peut rendre plus complexe la résolution du problème d'optimisation, ou encore aboutir à un équilibre instable. Il faudra alors s'assurer explicitement de ces propriétés.

On peut ainsi voir les fonctions d'utilité comme un compromis entre critère explicite et théorie des jeux. Dans le premier, on définit explicitement le critère d'ordonnancement instantané, sans en étudier l'impact sur une échelle temporelle plus longue qu'une étape d'ordonnancement. La théorie des jeux, elle, présuppose un comportement défini des utilisateurs, dont découle l'optimal. La définition de ce comportement est certes empirique, mais bien plus facilement justifiable : on dira qu'un flux élastique valorise le débit, alors qu'un flux temps réel sera plus attaché à la gigue et au délai. Cependant, le cadre mathématique ainsi défini est relativement lourd, pour un résultat final parfois équivalent aux critères d'ordonnancement. Les fonctions d'utilité présupposent deux hypothèses : les besoins des utilisateurs, représentés par la fonction d'utilité, et l'objectif du système, le problème d'optimisation. Les choix empiriques sont donc plus importants avec les fonctions d'utilité, mais ils permettent de donner plus de flexibilité aux algorithmes

d'ordonnancement et d'allocation de ressources. Certes, il est possible d'utiliser les fonctions d'utilité en Théorie des Jeux, où elles représentent le gain (ou revenu), mais il n'est pas assuré que cette formulation supplémentaire soit utile pour l'algorithme d'ordonnancement en lui-même, comme vu avec PF.

2.2.4 Conclusion de l'étude des outils analytiques

Nous avons proposé successivement trois outils analytiques pour l'ordonnancement et l'allocation de ressources : les critères explicites, la Théorie des Jeux, et enfin les fonctions d'utilité. Bien que liés, nous avons cherché à isoler, pour chaque outil, ses spécificités, ses avantages et inconvénients. Cette classification fait ressortir la difficulté de concilier simplicité (dans la conception aussi bien que dans l'implantation) avec des garanties apportées par un modèle analytique rigoureux. Sur le plan de la complexité algorithmique, nous pouvons remarquer que bien que la Théorie des Jeux et les fonctions d'utilité soient défavorisés, certains critères explicites peuvent être utilisés par des algorithmes tels que WFQ dont la complexité est comparable aux algorithmes de résolution de problème d'optimisation.

Nous avons également souligné la difficulté de lier la définition d'un optimal avec l'algorithme permettant de l'atteindre, dans le cas de la Théorie des Jeux comme dans celui des fonctions d'utilité. On peut donc s'interroger sur la possibilité d'atteindre effectivement cet optimal, et sur sa pertinence dans un système évoluant au cours du temps. Si l'ordonnancement *myope* semble être moins intéressant, quel gain est véritablement apporté par des méthodes anticipant l'impact des décisions d'ordonnancement ?

De plus, les frontières entre chaque famille d'outils ne sont pas strictes, d'autant moins si elles permettent d'obtenir un même résultat, comme pour PF. Nous pouvons cependant, à partir de cette description, analyser les différentes propositions d'ordonnancement sous un angle différent d'une simple évaluation de performances, notamment le rapport entre le critère d'ordonnancement, qui détermine la décision à chaque instant d'ordonnancement, et l'objectif d'ordonnancement, à plus long terme, qui décrit un comportement moyen du système. Nous avons à présent les outils nécessaires pour analyser les solutions d'ordonnancement et d'allocation de ressources pour les systèmes de Télécommunications par Satellite, proposées dans la littérature sur le sujet.

2.3 Revue des méthodes d’ordonnancement et d’allocation de ressources appliquées au Satellite

Dans cette partie, nous concentrons notre propos sur les propositions dans les domaines de l’ordonnancement (voie Aller) et de l’allocation de ressources (voie Retour) appliquées spécifiquement au Satellite. Sans objectif d’exhaustivité, nous retenons les solutions qui présentent des innovations certaines, ou constituent des travaux de référence. La voie Aller puis la voie Retour sont présentées, bien que certains travaux ne fassent pas cette distinction. Elle nous semble nécessaire compte tenu des différences exposées au Chapitre 1. La classification proposée dans la partie précédente sera mise à profit afin de comprendre les mécanismes analytiques présents dans chaque solution.

2.3.1 Voie Aller

La voie Aller a suscité un nombre important de travaux [4] [5] [61] [6] [34] [8] [7], car l’intérêt à optimiser cette voie de communication est grand : étant donné le volume du trafic acheminé par cette interface, une simple amélioration peut présenter un gain significatif en termes de rapport prix/performance.

Critères explicites

Un nombre important de propositions se concentre sur l’optimisation de l’efficacité spectrale, en particulier avec l’introduction de l’ACM dans DVB-S2. Dans [4], un ordonnanceur reposant sur un algorithme de WRR est présenté, conformément aux recommandations présentées dans [10]. Les paquets IP provenant du réseau terrestre sont d’abord encapsulés dans MPE. Ils sont envoyés dans la file d’attente correspondant au MODCOD du Terminal auquel ils sont destinés, comme indiqué sur la figure 2.2, et servis à l’aide d’un WRR. Cette architecture correspond à un ordonnanceur de paquets MPEG2-TS.

L’objectif de cet ordonnanceur est de fournir, avec un certain paramétrage, un débit utile égal pour chaque MODCOD, à savoir :

$$\forall m_1, m_2 R_{m_2} = R_{m_1} \tag{2.10}$$

Où R_{m_1} est le débit utile moyen du MODCOD m_1 (respectivement m_2). Cette condition permettrait d’obtenir un ordonnancement équitable entre chaque MODCOD, l’efficacité spectrale n’ayant alors plus d’impact dans l’ordonnancement. Les poids du WRR, c’est-à-

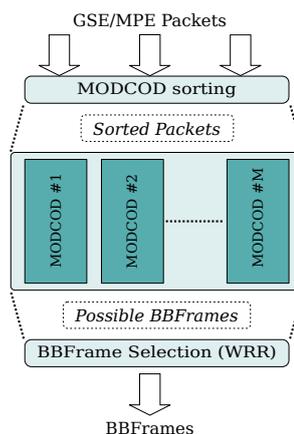


FIGURE 2.2 Architecture simple d'ordonnanceur

dire la proportion du temps consacrée à chaque file, sont définis par :

$$x_m = \frac{\frac{1}{T_m \eta_m^\alpha}}{\sum_{i=1}^M \frac{1}{T_i \eta_i^\alpha}} \quad (2.11)$$

Où :

- T_m est la durée d'une BBFRAME de MODCOD m ;
- η_m est l'efficacité spectrale associée au MODCOD m ;
- α est un paramètre d'équité, compris entre 0 et 1.

Les auteurs montrent dans [4] qu'une valeur $\alpha = 1$ permet d'atteindre l'objectif fixé en (2.10).

Cet algorithme se classe clairement dans la catégorie des critères explicites, puisqu'il propose une formulation empirique pour l'ordonnancement. L'objectif à long terme défini en (2.10) semble fournir une certaine garantie, mais il est plus adapté à un trafic régulier qu'à un trafic tel que décrit dans le Chapitre 1, du fait non seulement de la taille variable des paquets, incompatible avec le WRR¹, mais aussi du fait de la distribution inégale de la charge entre les MODCOD. En effet, le débit demandé sur chaque MODCOD dépend des conditions météorologiques des Terminaux, il est donc peu probable que chaque MODCOD ait besoin d'un débit équivalent, et qui plus est que ce débit soit stable dans le temps. Il est certes possible d'adapter α à la distribution du trafic, mais cette modification peut se révéler très complexe à mettre en place : il faudra introduire dans l'algorithme une boucle de contrôle sur α . Une différenciation de QoS est également proposée, *via* une modification du poids original en fonction de la QoS, ce poids étant assimilable à une

1. Des paquets de taille constante sont nécessaires pour prédire le comportement du WRR.

priorité.

Cet algorithme propose donc un algorithme d'ordonnancement simple, à l'implantation aisée et de faible complexité algorithmique, grâce au WRR. Il fait figure de référence, et de nombreuses améliorations ont été proposées [62],[63],[7],[64]. Cependant, son manque de flexibilité semble être un obstacle difficile à contourner, étant inhérent à sa conception. De plus, la prise en compte des contraintes de QOS par une seule priorité fixe est mal adaptée aux contraintes du trafic considéré.

Dans [6], les auteurs étudient la performance d'un ordonnanceur DVB-S2/GSE, avec un trafic entrant de type IP, et se rapproche donc plus de notre système de référence. Une fois encore, les paquets sont triés par MODCOD, avec une file d'attente pour chacun d'entre eux, et la charge globale est stable, avec une distribution de MODCOD de type ciel clair.

Les auteurs définissent cinq critères explicites d'ordonnancement :

- longueur de la file en nombre de paquets PQ ;
- longueur de la file en bits BT ;
- taux de remplissage de la BBFRAME PP ;
- temps de traitement restant, Longest Remaining Processing Time LRPT ;
- efficacité spectrale M.

Ces critères ont un fonctionnement très proche des règles d'ordonnancement. Le temps de traitement résiduel est le temps qu'il faudrait pour vider la file d'attente, au débit du MODCOD lui correspondant (et qui sera donc dépendant de l'efficacité spectrale). On distingue trois types de critères ici : les trois premiers sont strictement dynamiques, le dernier est statique (on choisit la BBFRAME pleine avec le meilleur MODCOD), et le quatrième représente un compromis entre ces deux types de critère. Les enseignements tirés de cette étude sont les suivants :

1. Les critères statiques (M) sont généralement inéquitables : l'efficacité spectrale seule ayant fortement tendance à privilégier les meilleurs MODCOD, laissant peu de ressources aux autres. C'est cependant avec ce critère que le débit est maximal.
2. Les critères entièrement dynamiques (PQ, BT, PP) présentent une bien meilleure équité, mais sont parfois inefficaces car ils ne considèrent pas (ici) l'efficacité spectrale, et donc l'efficacité de la transmission. Plus particulièrement, si PQ souffre de l'absence de prise en compte de la taille des paquets, BT et PP ne permettent pas d'obtenir de meilleur taux de remplissage des trames, malgré leurs indications plus fines.
3. Le temps de traitement résiduel (LRPT) offre les meilleures performances, le compromis entre efficacité spectrale et un critère dynamique (la longueur de la

file associée) semble être pertinent.

Nous remarquons ici une claire amélioration dans la flexibilité de l'algorithme d'ordonnancement comparativement à [4], mais elle ne donne pas d'indications quant au traitement de la qualité de service. L'usage de GSE permet ici un traitement efficace des paquets IP (ou de taille variable), notamment grâce à la fragmentation, ce qui conforte les choix de notre système de référence.

Des travaux de même nature figurent dans [5] et [61], où un algorithme appelé Urgency-Based Maximum Throughput (UBMT) est présenté. L'architecture de l'ordonnanceur est ici modifiée, puisqu'une isolation de chaque flux IP dans une file d'attente spécifique est proposée. Il y aura donc potentiellement plusieurs files ayant un même MODCOD, et leur nombre sera amené à évoluer avec le temps, au rythme des communications de niveau supérieur. Le modèle de trafic adopté comporte deux classes : le trafic Constant Bit Rate (CBR) et File Transfer Protocol (FTP). Ces deux classes ont des besoins de QoS différents : l'un ayant besoin d'un délai minimal, l'autre d'un débit aussi important que possible. Les tailles de paquets sont supposées fixes, bien que GSE soit utilisé pour l'encapsulation.

L'algorithme proposé est constitué de deux étapes : la sélection du MODCOD m^* et du nombre de paquets de chaque classe n_{cbr} et n_{ftp} , et celle des paquets effectifs de chaque classe.

La sélection du MODCOD se fait selon les deux possibilités suivantes

1. S'il existe des paquets CBR dont le délai dépasse le seuil autorisé à l'instant d'ordonnancement courant, ils doivent être envoyés. Ces paquets étant associés à un MODCOD, on construira autant de BBFRAME que de MODCOD nécessaires, la BBFRAME sélectionnée étant celle permettant d'encapsuler le plus de paquets CBR possible, ce qui détermine n_{cbr} . On détermine alors n_{ftp} afin de maximiser le débit de la BBFRAME. C'est la partie Urgency-Based (UB) de l'algorithme.
2. Si aucun paquet CBR ne voit son délai dépasser le seuil, on tente alors de construire la BBFRAME ayant le débit maximal. C'est la partie Maximum Throughput (MT) de l'algorithme.

Une fois l'une de ces étapes effectuée, m^* , n_{cbr} et n_{ftp} sont déterminés. Les paquets effectivement encapsulés sont choisis entre toutes les files du MODCOD sélectionné selon deux algorithmes : EDF pour les paquets CBR, et PF pour les paquets FTP, tous deux présentés précédemment.

Nous retrouvons donc ici l'utilisation de règles d'ordonnancement, bien que couplées à un mécanisme permettant de déterminer le MODCOD optimal, ainsi qu'une classification

par flux. Cet algorithme offre une prise en charge explicite des besoins de QoS, et ne suppose pas d'adaptation au scénario. On peut regretter la limitation à des paquets CBR et FTP de taille constante, qui affaiblit la pertinence de l'algorithme pour une utilisation avec un trafic Internet tel que décrit au Chapitre 1. De plus, l'étape de choix de m^* , n_{cbr} et n_{ftp} présenterait une complexité exponentielle, puisque le problème est équivalent au problème du sac à dos (ou *Knapsack*), qui est NP-complet [65]. Enfin, l'utilisation de règles d'ordonnancement sur une classification par flux peut se révéler particulièrement complexe : si le système présenté dans [5] comportait quelques dizaines de flux, tout au plus, notre système comporte plusieurs milliers de Terminaux, et au moins autant de flux : le système passera difficilement à l'échelle.

Nous pouvons citer également les travaux de [34] (délai, taille de la file d'attente), où un critère complexe d'ordonnancement, reposant sur de multiples statistiques, est présenté. Cet ordonnanceur comporte également deux étapes, une version modifiée de PF pour ordonner au niveau MODCOD, et un critère explicite entre les QoS d'un même MODCOD. Du fait de la complexité du critère utilisé, cette solution présente un intérêt moindre, d'autant qu'il est difficile d'évaluer l'influence du paramétrage de cette solution sans avoir recours à un nombre important de simulations.

Une autre solution, développée dans [8], prend appui sur l'observation de l'impact des phénomènes météorologiques sur le débit, par l'intermédiaire d'une estimation de la réduction de débit à courte échelle. Un WRR est également utilisé, de manière similaire à [4], mais cette fois au niveau IP. L'ordonnanceur de niveau BBFRAME n'est pas explicité, même si les auteurs recommandent les critères définis dans [6].

De manière générale, cette famille d'ordonnanceurs partage deux caractéristiques importantes. La première est l'approche empirique, liée aux critères explicites. On cherche avant tout à proposer directement une solution au problème de construction de la prochaine BBFRAME, en posant un compromis entre efficacité spectrale ou débit, et QoS, lorsqu'elle est présente. La deuxième caractéristique est la grande dépendance des solutions proposées vis-à-vis du scénario, qui se manifeste souvent dans les approximations faites sur la nature du trafic entrant (paquets de tailles fixes, demande uniforme, nombre d'utilisateurs). Si elles ne sont pas réhilitoires, elles ne permettent pas d'avoir une vision claire des performances de ces systèmes dans le scénario de référence que nous avons pris (QoS, tailles variables). De plus, si les premières solutions semblent relativement disparates, un compromis autour des règles d'ordonnancement commence à apparaître [5],[34], il serait alors intéressant de déterminer si une adaptation plus poussée est possible. L'intérêt, outre la réutilisation des travaux déjà effectués pour les systèmes terrestres, serait d'uniformiser les propositions d'ordonnanceurs empiriques pour la voie Aller.

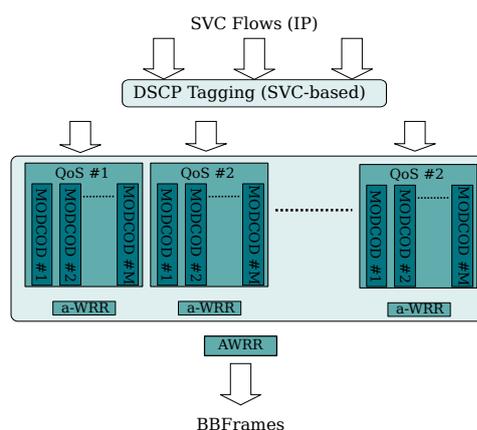


FIGURE 2.3 Architecture d'ordonnanceur proposée dans [7]

Fonctions d'utilité

Approche de Pradas Les fonction d'utilité sont au cœur des travaux de Pradas, exposés dans [66],[7] et [13]. Ces travaux sont les premiers à introduire l'utilisation des fonctions d'utilité pour l'ordonnancement dans les Systèmes de Télécommunications par Satellite, en particulier sur la voie Aller. Bien que l'objectif final, à savoir l'ordonnancement de plusieurs flux vidéo de type Scalable Video Coding (SVC), soit différent du notre, la méthode utilisée reste tout à fait pertinente. De façon schématique, SVC divise une vidéo de très haute définition en plusieurs flux additifs : le premier flux n'apporte qu'une vidéo de basse qualité, mais utilise peu de bande passante et sera robuste aux erreurs. Les flux supplémentaires nécessitent un débit plus important, mais, une fois additionnés au premier flux, permettent d'augmenter la qualité de la vidéo [67]. L'architecture proposée dans [7] assimile chaque code du SVC, soit un flux vidéo, à une QoS différente, à l'aide d'une classification de type DiffServ [68].

L'architecture simplifiée de cet ordonnanceur est donnée à la figure 2.3 : les paquets sont d'abord classés par QoS, c'est-à-dire par qualité de codage vidéo. Ils sont ensuite placés dans une file d'attente correspondant au MODCOD du Terminal auquel ils sont destinés. On trouve ici deux ordonnanceurs, l'un inter-QoS et l'autre intra-QoS, de manière similaire à [34].

Le premier ordonnanceur, inter-QoS, a pour objectif d'ordonner différents flux de même QoS mais ayant un MODCOD associé différent. On retrouve ici l'ordonnanceur présenté dans [4], qui est un critère explicite utilisant l'efficacité spectrale des MODCOD, l'expression du poids w_m étant donnée par (2.11). Ce choix est justifié ici par l'uniformité du trafic : vidéo de même niveau de codage, ce qui implique des paquets très similaires,

tant pour leur taille que leur débit d'arrivée. Cet ordonnanceur, reposant donc sur un WRR, est entièrement statique, et ne permet pas de prendre en compte une potentielle variation de la charge entrante, et souffrira des mêmes travers que [4].

Le cas du deuxième ordonnanceur est plus complexe, puisqu'il contient un problème d'optimisation découlant d'une définition à l'aide de fonctions d'utilité. Bien que la taxonomie de Théorie des Jeux (Joueurs, gain, etc. . .) soit mentionnée dans [7], l'absence d'utilisation concrète de résultats connus de Théorie des Jeux nous pousse à classer ces travaux dans la catégorie des seules fonctions d'utilité. Cet ordonnanceur est un WRR adaptatif (AWRR), c'est-à-dire dont les poids s'adaptent à l'état du système.

Plus précisément, les poids du WRR s'adaptent en fonction du délai moyen mesuré sur chacune des QOS, selon un ordre de priorité stricte entre QOS : en utilisant la dénomination DiffServ, EF est plus prioritaire que AF, lui-même dominant BE. Les poids $q_c(n)$ du WRR vérifient la relation suivante :

$$\forall n \sum_c q_c(n) = 1$$

Afin de fournir une garantie de délai à EF, son poids est augmenté autant que nécessaire, entraînant par conséquent une baisse du poids de BE, ou bien de AF si le premier n'est pas suffisant. Deux paliers de délai sont définis : au-delà du premier, c'est une partie du poids de BE qui est transférée à EF, selon une augmentation logarithmique. Si le délai dépasse un second seuil, c'est le poids de AF qui est mis à contribution.

Ces deux ordonnanceurs, en soi, n'utilisent pas les fonctions d'utilité, puisque nous observons deux WRR dont les poids sont définis selon des critères empiriques, w_m dépendant de l'efficacité spectrale et q_c de la QOS, ainsi que l'adéquation du service reçu aux besoins de la classe. On peut regretter ici la séparation en deux étapes de l'ordonnancement, tout comme [34] et [4], qui ne permet pas rendre compte clairement du compromis entre efficacité spectrale et QOS, ou encore entre capacité et demande, tels que définis au chapitre précédent.

Le véritable apport de [7] quant aux fonctions d'utilité se situe dans l'allocation de débit. Le but de la formulation est de déterminer le débit σ_c engendré par chaque flux SVC, sachant ses contraintes de QOS, ainsi que les Terminaux qui utilisent ce flux. En considérant chaque flux SVC comme une source indépendante, de fonction d'utilité :

$$U_c(\sigma_c) = \Omega_c \ln(\sigma_c)$$

les auteurs de [7] proposent une formulation sous forme de maximisation de l'utilité du

système, très proche de (2.6) :

$$\max_{\sigma_c} \sum_c U_c(\sigma_c)$$

avec une contrainte de capacité :

$$\sum_c \sigma_c \leq C$$

Les auteurs montrent que le débit σ_c est proportionnel à la priorité Ω_c , divisée par une constante dite d'échelle, indépendante de C , qui permet d'assurer la contrainte de capacité. Bien que connu, ce résultat était jusqu'ici absent de la littérature sur le satellite. L'innovation apportée par Pradas consiste à relier la priorité Ω_c au poids du WRR intra-qos q_c , menant ainsi à la relation suivante :

$$\Omega_c = (q_c)^{\delta_c} \propto \sigma_c \quad (2.12)$$

Où δ_c est un paramètre ajustable. On peut ainsi adapter le débit engendré par la classe c au poids du WRR correspondant.

Ainsi, les travaux de Pradas ont introduit le recours à des fonctions d'utilité dans les télécommunications par satellite par le biais d'une allocation en débit entre plusieurs flux vidéos. Bien que la formulation utilisée ne concerne pas l'ordonnancement à proprement parler, qui est laissé au WRR, la relation (2.12) permet de relier l'ordonnancement au débit des flux SVC.

On peut regretter cependant l'absence de considérations de dynamique temporelle dans la formulation du problème : outre le fait que la capacité exprimée en débit C est susceptible de changer au cours du temps, il semble difficile d'établir un lien, à l'échelle de l'ordonnancement, entre le débit applicatif σ_c et l'évolution temporelle du débit effectif délivré par le WRR, même si ces valeurs sont égales en moyenne.

On peut également remarquer que le trafic considéré (vidéo) possède de bonnes propriétés pour l'ordonnancement, comme vu dans le chapitre précédent. Ainsi, l'utilisation d'algorithmes simples, tels le WRR, est pertinente, mais ne permet pas d'envisager la présence de trafics de nature très différente. Or, c'est principalement pour ce type de scénario que les problèmes d'ordonnancement vont se poser, du fait du caractère apériodique du trafic, et peu prédictible sur une échelle de temps courte. On peut également souligner l'absence de considération sur l'ordonnancement au niveau BB-FRAME, et sur la gigue introduite par celui-ci, qui peut être préjudiciable pour un trafic vidéo. Enfin, la formulation proposée permet de calculer le débit σ_c indépendamment de sa nature : on pourrait parler ici de partage de capacité plutôt que d'ordonnancement. Comment envisager alors le cas où le WRR, n'ait aucun contrôle sur ce débit ? Cette

supposition est certes en contradiction avec le présupposé de [7], où l'application possède une utilité logarithmique, mais cette utilité correspond mieux à un trafic de type élastique, comme montré dans [46] plutôt qu'à un trafic vidéo, qui sera plutôt inélastique [69].

En guise de conclusion sur ces travaux, nous pouvons admettre l'intérêt de la méthode employée pour modéliser le problème d'ordonnancement, aussi bien pour le WRR que pour le partage de capacité. Cependant, l'application de cette méthode semble s'éloigner de notre scénario de référence, plus général. Elle souligne cependant l'intérêt d'étudier attentivement la composition du scénario afin de proposer une méthode adaptée et performante.

Approche de type Knapsack Dans [28], Chaput décrit un ordonnanceur pour DVB-S2/GSE, à l'aide d'une formulation sous forme de problème d'optimisation de fonctions d'utilité. Cet ordonnanceur ne présuppose aucune caractéristique particulière sur le trafic entrant, si ce n'est la présence de différentes contraintes de QoS, qui sont prises en compte dans la fonction d'utilité associée à chaque utilisateur. Contrairement à [7], les sources de trafic ne sont pas clairement identifiées : seuls les flux GSE sont considérés, sachant que ces derniers peuvent correspondre à une transmission unicast ou broadcast, transportant des applications diverses (vidéo, VOIP, FTP, etc. . .). Chaque flux GSE possède une file d'attente séparée, associée à un MODCOD optimal, de la même manière que l'architecture retenue dans le chapitre précédent.

On y retrouve une formulation très similaire à (2.6), avec une utilité en logarithme du débit moyen mesuré. L'originalité de [28] est de proposer une résolution sous-optimale de ce problème, amenant ainsi à une expression de PF adaptée à l'ordonnancement DVB-S2/GSE. La BBFRAME optimale est celle permettant de résoudre le problème suivant :

$$\max \sum_i \frac{l_i}{\bar{r}_i(n)\tau_m} \quad (2.13)$$

Sous la contrainte suivante :

$$\sum_i l_i \leq C_m$$

Où :

- l_i est le volume servi dans la BBFRAME pour le flux i ;
- τ_m est la durée de la BBFRAME de MODCOD m (durée dépendante de l'ordre de la modulation) ;
- $\bar{r}_i(n)$ est le débit moyen servi mesuré pour le flux i à l'instant n ;
- C_m est la taille utile de la BBFRAME de MODCOD m (dépendante du taux de

codage).

À la différence de (2.3), le critère d'ordonnancement se présente sous la forme d'un problème d'optimisation, dont les inconnues sont multiples :

- Quels flux(x) GSE faut-il servir (sur quels indices i la somme se fait-elle) ?
- Quel MODCOD m faut-il choisir ?
- Quel volume l_i faut-il choisir ?

Ce problème est en fait complexe : comme nous l'avons vu précédemment, la première et la deuxième inconnue sont inter-dépendantes, puisque un flux GSE est associé à un MODCOD. De plus, la variable l_i cache une complexité supplémentaire, dans le fait que celle-ci est le résultat d'une somme de tailles de paquets. En effet, bien qu'il soit possible avec GSE de considérer un volume continu, nous avons en réalité affaire à un ensemble de paquets. De ce fait, considérer un volume continu et fragmenter systématiquement peut mener à une efficacité réduite, du fait de l'overhead induit par la fragmentation.

Par conséquent, nous pouvons assimiler (2.13) à un problème de Knapsack (ou sac à dos), d'autant plus complexe que la capacité du sac, C_m est variable. La résolution exhaustive de ce problème est bien plus complexe que celle proposée dans [5] et [61], où le nombre de cas à traiter est de quelques dizaines. Ici, l'absence de présupposé sur la taille des paquets, ainsi que le nombre de flux rend la résolution exhaustive presque impossible, sinon beaucoup trop complexe pour notre contexte.

Une résolution sous-optimale est néanmoins proposée dans [28], prenant appui sur l'observation suivante : si le paquet en tête de file n'est pas servi, alors les suivants ne peuvent l'être, ce qui permet d'élaguer largement l'arbre de décision. Reste alors à déterminer si le paquet en tête de file doit être traité ou non, ce qui reste un problème complexe en soi.

La simplicité de la formulation de [28] est un avantage certain, et la formulation proposée possède les avantages de PF présentés plus haut : définition d'un optimal, ordonnancement non myope, mesure du gain de chaque BBFRAME et prise de décision d'ordonnancement en conséquence. Ces travaux permettent en résumé d'introduire l'ordonnancement directement à l'aide de fonctions d'utilité dans le contexte satellite. Cependant, la complexité rédhibitoire de l'algorithme lui-même ne permet pas d'obtenir une évaluation aussi complète que celle présentée dans [13], et l'implantation de l'algorithme dans un contexte temps réel semble compromise, réduisant la pertinence de la solution proposée, malgré un avantage certain quant aux performances.

Cette proposition constitue donc une base de travail pertinente, mais à laquelle il faudra apporter des réponses quant à la complexité de mise en œuvre, la prise en charge de la QoS, ainsi que l'implantation de l'algorithme lui-même.

Théorie des Jeux

Mis à part la dénomination de Théorie des Jeux présente dans les travaux de [13], cet outil est absent de la littérature sur l'ordonnancement voie Aller dans les systèmes satellite. On peut expliquer cette absence par plusieurs facteurs, le principal étant la difficulté de définition des Joueurs dans le contexte système. En effet, la Théorie des Jeux montre son intérêt principalement pour modéliser des situations de compétition où plusieurs entités peuvent prendre des décisions de manière indépendante, introduisant une notion sous-jacente de système distribué. Or le système que nous étudions est très centralisé : une seule entité (l'ordonnanceur) prendra la décision d'ordonnancement. Il est certes possible de considérer comment ces mêmes sources vont *réagir* à cet ordonnancement, mais cette étude suppose des considérations allant bien au-delà du cadre système que nous nous sommes fixés, limité à l'ordonnancement au niveau accès. Du reste, ces considérations n'ont rien de spécifique au Satellite, et constituent un thème de recherche à part entière [45], hors du contexte de cette étude.

2.3.2 La voie Retour

La voie Retour présente une diversité bien moindre des propositions pour l'allocation de ressources que la voie Aller pour l'ordonnancement. On peut l'expliquer tout d'abord par l'intérêt plus discutable de méthodes très efficaces sur la voie Retour, lorsqu'elles sont absentes de la voie Aller, où le trafic, et donc l'impact, est bien plus important. Étant donné que le lien dimensionnant d'un système est la voie Aller, il est souvent bien plus bénéfique pour l'ensemble du système d'optimiser l'ordonnancement de la voie Aller. De plus, la nature statique de la trame DVB-RCS laissait peu de marge de manœuvre à l'allocation de ressources, le problème se situant dans le dimensionnement de la Trame, plus que dans l'allocation des timeslots dont elle est constituée. DVB-RCS2 étant un standard relativement récent (2012), peu de travaux se sont déjà attelés aux problèmes spécifiques à cette nouvelle version, ce qui est précisément notre propos.

Deux axes de travail s'offrent alors à nous : étudier le calcul des requêtes, et l'allocation résultante sous un algorithme de référence, ou bien étudier l'allocation de ressources en prenant un calcul des requêtes de référence. C'est la deuxième solution qui nous intéressera ici, car le processus d'allocation des ressources en DVB-RCS2 est un problème nouveau, relativement peu traité. De plus, l'étude faite dans [70] montre clairement la difficulté d'obtenir de bons résultats, même avec des méthodes avancées telles que le prédicteur de Smith.

Critère explicite

La notion de critère explicite appliquée à l'allocation de ressources en DVB-RCS2 signifie qu'une simple fonction des paramètres du système permet de déterminer cette allocation, fonction définie empiriquement. Un exemple caractéristique de ce type d'algorithme est présenté dans [71], qui est par ailleurs plus centré sur l'estimation correcte des requêtes que sur l'allocation de ressources en elle-même. Les auteurs de [71] considèrent les requêtes de trafic exactes, c'est-à-dire correspondant effectivement à ce dont le Terminal aura besoin. Les requêtes sont alors traitées, à chaque Supertrame, selon un ordre de priorité stricte : d'abord Constant Rate Assignment (CRA), ensuite Rate-Based Dynamic Capacity (RBDC) puis Volume-Based Dynamic Capacity (VBDC). Ceci permet de simplifier grandement l'allocation de ressources, puisque seuls deux cas sont à traiter : la sous-réservation, lorsque la somme des requêtes pour une Supertrame donnée est inférieure à sa capacité (le nombre de timeslots), ou bien la sur-réservation, lorsque cette somme est supérieure à la capacité. Le problème de la sous-réservation est traité avec le Free Capacity Assignment (FCA), qui redistribue la capacité non utilisée de façon à maximiser l'efficacité spectrale. Le cas de la sur-réservation est traité en appliquant à chaque requête un coefficient, dépendant de la requête, permettant de garantir que l'allocation ne dépasse pas la capacité. Formellement, si C est la capacité allouée sur une Supertrame (en retirant le CRA déjà alloué), et x_i la requête RBDC du Terminal i , l'allocation sera alors :

$$alloc_i = \frac{x_i}{\sum_j x_j} C \quad (2.14)$$

Cette allocation reste très simple, et ne permet pas une prise en compte de la QoS, ni des conditions de transmission, ni de l'évolution de la capacité. Elle permet néanmoins de répondre au problème d'allocation, en supposant des requêtes de trafic correctement calculées, ce qui n'est pas nécessairement trivial. De plus, aucune métrique moyenne n'apparaît ici, du fait des contraintes du système : les seules informations venant des Terminaux disponibles pour l'allocation de ressources sont les requêtes de trafic, et éventuellement les MODCOD utilisés. La Passerelle pourrait certes mesurer elle-même le débit atteint par un Terminal, ou bien l'adéquation entre la requête et l'allocation, ou encore entre l'allocation et les ressources effectivement utilisées, mais ces informations ne sont pas nécessaires au fonctionnement de l'algorithme, et n'apportent qu'une information *a posteriori*, très retardée à cause du temps de propagation.

On peut noter également que, bien que ces travaux soient adaptés à DVB-RCS, ils sont applicables à DVB-RCS2, en admettant quelques modifications sur les contraintes : le format de la Supertrame étant cette fois dynamique, et DVB-RCS2 utilise plusieurs

modulations là où DVB-RCS n'utilise que QPSK.

Fonctions d'utilité

Le domaine satellite ne compte pas de travaux sur l'allocation de ressources mentionnant seulement les fonctions d'utilité comme outil analytique, du moins en ce qui concerne les travaux sur DVB-RCS et DVB-RCS2. Cependant, l'approche consistant à définir l'allocation comme solution d'un problème d'optimisation est une technique largement utilisée dans le domaine. Nous examinerons dans cette partie deux propositions, [72] et [73].

Dans [72], les auteurs proposent un algorithme d'allocation de ressources interactif, reposant sur la différenciation entre le trafic temps réel (CBR) et BE. La solution proposée mélange le CAC pour la gestion des flux temps réel (des flux de VoIP) et un algorithme d'allocation de ressources prenant en compte le taux de perte pour BE. Opérant Supertrame par Supertrame, tout comme (2.14), l'algorithme procède d'abord à l'allocation des flux CBR, ce qui permet de calculer la capacité résiduelle C , disponible pour les flux BE. L'allocation Z_i donnée à chaque Terminal doit être solution du problème d'optimisation suivant² :

$$\min \sum_i \bar{P}_{loss}(Z_i)$$

Où \bar{P}_{loss} est l'estimation du taux de pertes moyen sur tous les flux du Terminal i , sachant une allocation égale à Z_i (également dépendante du MODCOD et de la capacité restante). On retrouve une contrainte de capacité classique :

$$\sum_i Z_i \leq C$$

Le comportement d'une telle formulation est évidemment très dépendant de l'expression de \bar{P}_{loss} , et donc du modèle de trafic utilisé. La méthode est cependant intéressante, car elle correspond à celle des fonctions d'utilité, en considérant \bar{P}_{loss} comme la fonction d'utilité, et Z_i comme la métrique. On note qu'il aurait été possible de prendre également en compte le taux de perte maximum, permettant une garantie de QoS supplémentaire.

On peut remarquer à la lecture de [72] combien cette solution est cependant dépendante du système, notamment du format de la trame DVB-RCS, du fait qu'une seule modulation est disponible, QPSK, et surtout de l'absence de prise en compte des requêtes. En effet, l'allocation pour les flux BE pourrait profiter de requêtes RBDC ou VBDC afin d'obtenir une meilleure efficacité de l'allocation, mais cette option n'est pas considérée par [72]. De

2. Donnée ici dans une version simplifiée

plus, la mesure du taux de perte moyen *a posteriori* semble adaptée pour des flux longs, ce qui n'est pas le cas de la plupart des flux BE, exceptées les connexions FTP [74].

Cette formulation est en partie reprise dans [73], où un algorithme permettant d'optimiser la Trame DVB-RCS2 est présenté. Ces travaux sont – du moins à notre connaissance – les seuls concernant l'allocation de ressources citant explicitement DVB-RCS2. L'objectif défini dans [73] est cependant très proche d'un mode de fonctionnement de type DVB-RCS, puisqu'il propose un algorithme d'optimisation afin de définir le nombre de timeslots dédiés à chaque MODCOD (DRA dans le texte). La fonction objectif définie tente de faire un compromis entre efficacité spectrale et équité :

$$J(z) = w_{THR} \sum_d z_{ST}(d) + w_{FAIR} \min_d \frac{z_{ST}(d)}{REQ_{ST}(d)}$$

Où :

- $z_{ST}(d)$ est le nombre de timeslots réservés au MODCOD d (z est le vecteur des z_{ST}) ;
- $REQ_{ST}(d)$ est la requête totale des Terminaux ayant pour MODCOD d ;
- w_{THR} et w_{FAIR} sont des poids.

Si le premier terme de cette fonction optimise le débit, le second exprime l'adéquation entre ressources demandées et ressources disponibles. Cette fonction est définie pour deux types de trafic : un de haute priorité, l'autre de basse priorité, éventuellement complétés par une catégorie réservée au FCA, de manière très similaire à [72]. On tente alors de maximiser la somme des $J(z)$ pour chaque type de trafic, les auteurs de [73] montrant qu'on peut se ramener à un problème de programmation linéaire sous contraintes, pour lequel il existe de nombreux algorithmes efficaces.

Le point intéressant de cette proposition est l'échelle temporelle de résolution : si la plupart des algorithmes se contentent de résoudre Supertrame par Supertrame, les auteurs de [73] proposent de définir une période, T_{FRAMER} , comprenant plusieurs Supertrames. Le nombre de timeslots obtenus en résolvant le problème exposé plus haut devra donc s'étaler sur toute cette période, et donc sur plusieurs Supertrames. L'objectif est ici de s'adapter à la demande moyenne des Terminaux ainsi qu'au conditions météorologiques, qui changent relativement lentement par rapport à la période de la Supertrame. Les auteurs donnent une valeur d'environ 300s pour T_{FRAMER} . Il est également intéressant de noter que les auteurs ne donnent pas l'algorithme d'allocation des ressources en lui-même, processus explicitement séparé de l'optimisation du format de la Supertrame.

L'approche de [73] permet de définir de manière rigoureuse le format de la trame DVB-RCS2, mais reste très proche d'un mode de fonctionnement DVB-RCS, où le nombre

de timeslots par MODCOD est fixé. De plus, la séparation entre la définition du format de la Supertrame, et l'allocation des timeslots à proprement parler ne permet pas de tirer parti des innovations apportées par DVB-RCS2, en particulier la définition de la Supertrame à l'aide de BTU, qui apporte beaucoup plus de flexibilité que les timeslots DVB-RCS.

Théorie des Jeux

Le recours à la Théorie des Jeux sur la voie Retour semble être plus adapté que sur la voie Aller. En effet, il est aisé de définir les Terminaux comme autant de Joueurs devant se partager les ressources disponibles, la Passerelle pouvant éventuellement forcer ou non la coopération entre Terminaux. On trouve cette formulation dans trois travaux très similaires, [75], [63], et [13].

On remarque une relative homogénéité dans ces travaux, où le problème d'allocation des ressources est formulé à l'aide de l'Équilibre de Nash, qui donne lieu à un NBS. L'hypothèse d'une coopération entre Joueurs, nécessaire pour atteindre l'équilibre, est ici tout à fait plausible, et peut être assurée par l'intermédiaire de la Passerelle.

Le problème est généralement formulé comme une variante de (2.6), puisque seul le débit est considéré comme gain³. Sous des hypothèses facilement vérifiées de convexité, il est possible de résoudre le problème d'optimisation ainsi posé à l'aide de méthodes Lagrangiennes [76].

De manière générale, l'expression du débit moyen optimal pour le Terminal i est donnée par :

$$\bar{r}_i^* = \frac{p_i}{\lambda + c_i}$$

Où

- r_i^* est le débit optimal ;
- p_i est une priorité (par exemple de QOS) ;
- c_i est un coût fixé par l'opérateur ;
- λ est le multiplicateur de Lagrange, dont la valeur peut être calculée [63].

On peut faire pour ces travaux des remarques similaires à celles faites sur la voie Aller : le cadre mathématique permet certes un calcul précis de l'allocation en débit, c'est-à-dire de ce que les Terminaux devraient, dans l'idéal, demander ou obtenir. En revanche, cette allocation ne prend pas en compte les besoins effectifs des Terminaux, qui peuvent être bien différentes. De plus, cette allocation ne permet pas directement de donner l'allocation de ressources telle que nous l'avons définie dans le chapitre précédent.

3. Dans [13], cette formulation correspond au troisième problème

De plus, le facteur temporel est absent de la plupart des formulations, ce qui s'apparente une fois de plus à un partage de capacité plutôt qu'à la définition que nous avons prise. Ainsi, bien que ces méthodes soient d'un intérêt évident dans la conception d'un système, elles ne permettent pas, en l'état, de répondre à notre problème d'allocation de ressources.

Conclusion sur l'état de l'art

Il apparaît, après analyse des différentes propositions concernant l'allocation de ressources et l'ordonnement, qu'aucune solution ne permet véritablement de répondre de manière satisfaisante au problème posé dans le chapitre précédent, dont l'une des particularités est d'avoir des caractéristiques de trafic disparates.

En effet, sur la voie Aller, de très nombreuses propositions ont été faites, mais peu d'entre elles arrivent à couvrir l'ensemble des points identifiés comme clés, tels que la prise en compte de paquets de taille variable, et donc de GSE, la QoS et un trafic majoritairement IP, et l'ordonnement de niveau accès ou BBFRAME ayant une complexité acceptable pour un système réel. Des solutions prometteuses sont cependant disponibles, telles que les règles d'ordonnement, des critères explicites comme dans [5], ou encore l'ordonnement à l'aide de fonctions d'utilité de [28]. Nous nous appliquerons donc à proposer une solution permettant de répondre à l'ensemble de ces besoins.

Sur la voie Retour, la faible quantité de travaux spécifiques à DVB-RCS2 montre le travail qu'il reste à effectuer afin de pouvoir exploiter les atouts apportés par ce nouveau standard. De plus, le caractère très spécifique de la voie Retour satellite, avec le mécanisme des requêtes de capacité conjuguées à un délai de propagation très long, empêche une adaptation rapide des travaux menés dans le domaine terrestre. Nous concentrerons donc notre travail sur la proposition d'un algorithme de DAMA, prenant en compte la nature du trafic, ainsi que les spécificités de DVB-RCS2.

ALGORITHMES D'ORDONNANCEMENT POUR LA VOIE ALLER

Dans ce chapitre, nous présentons les solutions d'ordonnement pour la voie Aller d'un système de télécommunications par satellite offrant un accès de type broadband. Notre objectif est ici de proposer une réponse convaincante aux problèmes introduits dans les chapitres précédents, c'est-à-dire comment construire un ordonnanceur de BBFRAME, prenant en compte la QOS, le caractère variable de la capacité, ainsi que de la demande.

L'étude faite au Chapitre 2 nous a permis de dégager les axes de travail suivants. Tout d'abord, il est impératif de prendre en compte, dès la conception, les spécificités de DVB-S2 et de GSE, c'est-à-dire l'ACM, la prise en compte des paquets de taille variable, et l'encapsulation. Ensuite, la crédibilité de nos propositions sera mesurée par leur capacité à prendre en charge un trafic très disparate, possédant des contraintes de QOS hétérogènes. Cette notion devra donc également être intégrée dans notre conception. De même, le caractère dynamique de la capacité est une notion importante, qu'il faudra pouvoir formuler à l'échelle temporelle de l'ordonnement. Enfin, si notre objectif est de proposer une solution performante et de faible complexité, notre développement devra être poursuivi jusqu'à proposer une implantation de référence, dont la complexité soit compatible avec un ordonnement en temps réel.

Deux solutions semblent prometteuses face à ces objectifs : les règles d'ordonnement, en supposant qu'un bon paramétrage puisse être trouvé, et les fonctions d'utilité. La Théorie des Jeux pourrait être considérée, mais n'apporte pas d'avantage significatif, par rapport aux fonctions d'utilité, pour notre problème en particulier, en plus de la difficulté de modélisation de compétition entre plusieurs acteurs indépendants.

Nous présentons tout d'abord l'architecture de QOS retenue, ainsi que les algo-

algorithmes de référence utilisés afin de comparer les performances de notre solution avec les mécanismes existants. Nous présentons ensuite notre adaptation des règles d'ordonnancement à notre contexte, puis la formulation d'un ordonnancement à l'aide de fonctions d'utilité. Après une discussion sur les implantations possibles des solutions proposées, nous présentons une évaluation des performances comparées des solutions développées, et proposons une solution adaptée aux besoins du satellite. A travers ce chapitre, nous montrons que les solutions que nous avons développées permettent de répondre aux objectifs donnés précédemment.

3.1 Architecture de l'ordonnanceur

Comme nous avons pu le constater dans les chapitres précédents, l'architecture de l'ordonnanceur possède une influence significative sur l'algorithme d'ordonnancement. Par exemple, il est commun de classer les paquets par MODCOD, car cette solution simplifie largement la formation des BBFRAME. Nous reprenons donc cette classification dans l'architecture de notre ordonnanceur. Elle n'est cependant pas suffisante pour assurer une solution compatible avec nos objectifs, puisqu'elle ne permet pas la prise en charge de la QoS. Nous spécifions dans la suite l'architecture de QoS utilisée dans notre ordonnanceur de référence, ainsi que les files d'attente considérées.

3.1.1 Choix de la QoS

Si l'architecture de QoS se réduit, dans beaucoup de travaux sur la couche accès, à une simple priorité, il nous semble ici nécessaire de spécifier plus précisément le type de QoS que nous allons considérer. Le choix de cette architecture n'est pas indépendant des contraintes de notre système de référence, soit en l'occurrence la taille du système, la présence de plusieurs MODCOD et le type de trafic considéré.

Étant donné le nombre de Terminaux, il semble qu'utiliser une architecture de QoS isolant les Terminaux ne permettra pas de passer à l'échelle : il faudrait alors plusieurs milliers voire dizaines de milliers de files d'attente. On peut certes supposer une isolation des flux, au niveau IP de la Passerelle, et une gestion des files d'attente avancée. Cette considération est d'autant plus pertinente que la voie Aller constitue certainement un goulot d'étranglement, sujette aux phénomènes de *bufferbloat*[77]. Cependant, ces considérations ne concernent pas la couche accès en particulier, aussi supposons-nous que le multiplexage effectué au niveau de la couche accès, entre les différents flux à destination des Terminaux, est mis en œuvre de manière à éviter ces phénomènes.

Aussi, considérant les trois types de trafic définis dans le chapitre 1, proposons-nous

une architecture de QOS reposant sur Differentiated Services (DIFFSERV) [68]. Cette architecture permet une différenciation de QOS, tout en restant simple et indépendante du nombre de Terminaux. Notons également qu'elle suppose un marquage des paquets IP, qui ne peut être utilisé efficacement qu'à l'intérieur d'un réseau entièrement sous le contrôle d'un seul opérateur. DIFFSERV associe, pour chaque nœud du réseau, un traitement spécifique selon le marquage du paquet, appelé Per Hop Behaviour (PHB). Ce traitement est une définition locale, et non de bout-en-bout de la QOS, bien qu'il puisse être étendu à plusieurs nœuds du réseau. Ainsi, ce marquage DIFFSERV sera *a priori* réalisé de bout-en-bout pour le trafic VOIP et vidéo provenant du réseau de l'opérateur, mais seulement sur le réseau de l'opérateur pour le trafic BE. Cette limitation n'est cependant pas un réel problème, puisque le trafic BE ne dispose pas de garanties de QOS.

Nous utiliserons trois marques DIFFSERV (correspondant à trois classes de QOS), Expedited Forwarding (EF), Assured Forwarding (AF) et BE, correspondant respectivement au trafic VOIP, vidéo, et BE(Internet). Le traitement de chaque classe dépend de ses caractéristiques : EF doit être servi le plus régulièrement possible, avec un délai faible et constant, AF est un peu moins sensible au délai, mais tolère mal la gigue, tandis que BE demandera un débit potentiellement important sur la voie Aller.

3.1.2 Architecture proposée

L'architecture de l'ordonnanceur est présentée à la figure 3.1. Les paquets IP sont marqués en fonction de leur provenance, puis dirigés dans une file d'attente correspondant au MODCOD du Terminal auquel ils sont destinés. Ainsi, on trouvera autant de files d'attente d'une même QOS que de MODCOD. Si DVB-S2 spécifie jusqu'à une trentaine de MODCOD différents, l'expérience montre que seuls quelques-uns sont simultanément utilisés [28] [23], limitant le nombre total de files actives pour une classe de QOS à une dizaine, au maximum.

Les files d'attentes ainsi définies sont les utilisateurs de notre algorithme d'ordonnement, comme définis au Chapitre 2. On note ici une architecture faisant usage d'informations provenant de la couche IP, à savoir la marque DiffServ, ainsi que de la couche physique, à savoir le MODCOD. Nous sommes donc en présence d'une architecture dite «Cross-Layer», permettant un passage d'informations entre les couches. On paramètre ces files par un double indice, m pour le MODCOD et q pour la classe de QOS, et on posera dans la suite $(m, q) = k$ de sorte à ne garder qu'un seul indice dans les expressions. On prendra, par convention, l'indice $m = 0$ pour le plus faible MODCOD, et $m = M - 1$ pour le meilleur, M étant le nombre total de MODCOD. Pour la QOS, $q = 0$ correspondra à la QOS la plus haute, $q = Q - 1$ à la plus basse, Q étant le nombre de

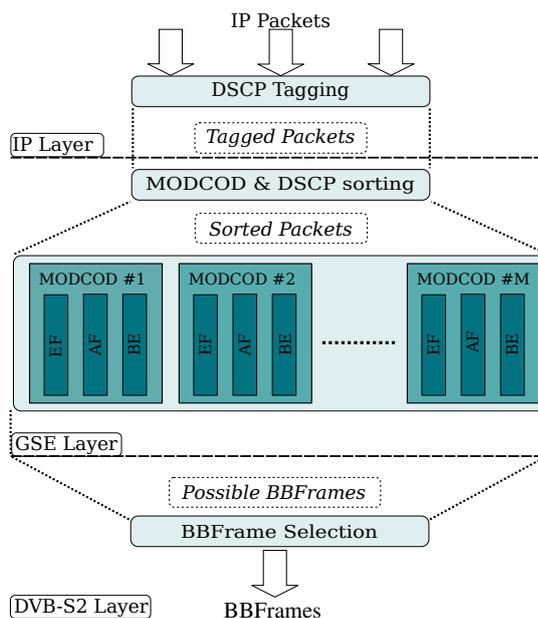


FIGURE 3.1 Architecture de l'ordonnanceur DVB-S2/GSE

QoS. Nous noterons $N = MQ$ le nombre total d'utilisateurs dans le système, définis ici comme files d'attente. Nous noterons ν_m l'ensemble des indices k désignant une file dont le MODCOD est au moins m , c'est-à-dire que ν_m comprend les utilisateurs du MODCOD m et tous ceux des MODCOD supérieurs. Cette notation permet de prendre en compte le déclassement dans les expressions, plutôt que les seuls utilisateurs du MODCOD m .

On peut noter que cette définition contraint la conception des fonctions d'utilité, puisqu'une file de ce type n'accordera pas de l'importance aux mêmes métriques qu'une file d'attente correspondant à un seul flux, ou à un Terminal. De même, cette file n'aura pas un comportement autonome, puisqu'elle se contente de multiplexer les trafics de différents flux, et sera donc tributaire de leur comportement.

Les différentes solutions d'ordonnancement seront ici constituées de paquets pris dans chaque file, formant une BBFRAME. La décision d'ordonnancement consistera à choisir entre ces BBFRAME la meilleure selon le critère défini par l'algorithme d'ordonnancement.

Dans les solutions d'algorithmes d'ordonnancement déjà présentés, nous retiendrons trois références : [5], [4], et [28] toutes trois compatibles avec notre modèle, au prix de quelques modifications, principalement afin d'adapter ces solutions à des paquets de taille variable. Pour [4], nous utiliserons un DRR à la place du WRR (qui ne prend pas en compte les paquets de taille variable), où le quantum est paramétré par les poids définis à l'équation (2.11). Pour [5], l'adaptation nous avons considéré une version simplifiée, où

seule la partie de sélection du MODCOD est présente, puisqu'il n'y a qu'une seule file par MODCOD, pour une QoS donnée. Ainsi, il suffira de déterminer la file dans laquelle le plus de paquets CBR dépassant leur délai maximal sont présents pour déterminer le MODCOD et une partie de la BBFRAME. Remarquons que cette solution suppose une recherche exhaustive sur toutes les files. La solution proposée dans [28] sera reprise dans notre proposition reposant sur les fonctions d'utilité.

3.2 Règles d'ordonnancement pour DVB-S2/GSE

Les règles d'ordonnancement permettent de disposer d'un cadre commun pour l'élaboration d'algorithmes d'ordonnancement dans les réseaux sans fil, tout en adoptant une démarche empirique. Leurs performances se révèlent généralement bonnes, pour une complexité faible, ce qui en fait une solution de choix. Cependant, les systèmes satellite présentent quelques différences avec leurs homologues terrestres, qu'il faudra prendre en compte dans la définition des règles d'ordonnancement pour DVB-S2/GSE.

Aussi, nous présentons tout d'abord les critères selon lesquels nous avons adapté les règles d'ordonnancement, et l'impact attendu de cette adaptation sur leur comportement. Nous avons retenu trois règles d'ordonnancement pour notre étude, PF, M-LWDF et EXP-PF, du fait de leur popularité. Une fois de plus, PF nous servira d'exemple de référence dans l'adaptation.

3.2.1 Principes de l'adaptation des règles d'ordonnancement

Les règles d'ordonnancement ont été définies dans un contexte de réseaux mobiles terrestres sans fil, tels que LTE. PF a connu un franc succès, du fait de sa simplicité et de sa performance [48], et joue le rôle d'un standard de fait, dont de nombreuses règles s'inspirent (M-LWDF, EXP-PF). Si notre objectif est de proposer une adaptation aux satellite de ces règles, PF pourra donc servir de référence.

PF définit un ordre de priorité des utilisateurs en fonction de leur débit moyen et de leur débit atteignable à cet instant d'ordonnancement. Dans les systèmes de type LTE, la périodicité de l'instant d'ordonnancement est de l'ordre de la milliseconde [25], et cette durée strictement constante, du fait de la nature de l'accès en OFDMA. Dans les exemples classiques donnés [6],[34] pour DVB-S2, la durée d'une BBFRAME est du même ordre de grandeur, il n'y a donc pas *a priori* d'adaptation sur l'échelle temporelle.

Si cet ordre de priorité peut être utilisé pour ordonnancer un certain nombre de ressources indépendantes, comme des Timeslots dans OFDMA, ce problème reste bien plus complexe lorsqu'il s'agit d'une BBFRAME. En effet, nous avons vu que la sélection

du MODCOD, qui définit la taille de la BBFRAME, contraint les paquets qui peuvent y être encapsulés. Ainsi, une simple sélection selon un critère PF entre toutes les files de l'architecture donnée à la figure 3.1 ne permettra pas de déterminer entièrement la prochaine BBFRAME, seulement une file parmi toutes.

Aussi, nous définissons deux règles : la règle utilisateur, et la règle MODCOD, qui seront formulées de façon à permettre conjointement le choix du MODCOD et des paquets composant la BBFRAME.

Règle utilisateur

La règle utilisateur permet d'ordonnancer au sein d'un même MODCOD. La règle de l'utilisateur k est notée $c_k(n)$. Dans notre architecture, un utilisateur est associé à un MODCOD optimal. Ses paquets peuvent être envoyés dans une BBFRAME ayant un MODCOD plus faible (*déclassement*) mais cette solution n'est pas optimale quant à l'efficacité spectrale, et ne doit être utilisée que pour combler un vide dans une BBFRAME. Son débit atteignable est donc fixé, et égal au débit du modcod, soit R_b tel que défini en (1.1). Contrairement aux réseaux sans fil terrestres, ce débit atteignable n'évolue pas pour un utilisateur de l'ordonnanceur défini. Cette différence ici est de l'ordre de l'architecture : si nous avons considéré une file par Terminal, le débit atteignable aurait suivi les évolutions du MODCOD associé au Terminal. Cependant, ce débit atteignable aurait changé lentement, en comparaison de l'horizon temporel de l'ordonnancement. Cette construction semble donc être en accord avec notre contexte.

On peut définir alors la règle utilisateur comme :

$$k^* = \arg \max_{k \in \nu_m} \delta_k(n) c_k(n) \quad (3.1)$$

Où ν_m , défini plus haut, est l'ensemble des utilisateurs ayant pour MODCOD au moins m , et $\delta_k(n)$ vaut 1 si des paquets sont présents dans la file k à l'instant d'ordonnancement n , zéro sinon. Ce dernier paramètre permet d'éviter qu'une file vide ait une influence dans la décision d'ordonnancement. Comme nous l'avons déjà indiqué dans le Chapitre 2, cette formulation ne définit pas un algorithme en soi, il faudra encore spécifier comment cette règle est utilisée. Cette règle aura une définition similaire à celles déjà présentées, excepté qu'elle ne contiendra pas de débit atteignable.

Dans le cas de PF, la règle utilisateur sera définie de la façon suivante :

$$c_k(n) = \frac{1}{\bar{r}_k(n)} \quad (3.2)$$

Où $\bar{r}_k(n)$ est le débit moyen servi de la file k . En supposant que le MODCOD m a été sélectionné pour la prochaine BBFRAME, nous pouvons donc sélectionner les paquets la composant selon la règle (3.2), qui va privilégier les utilisateurs dont le débit servi moyen est faible.

Règle MODCOD

La règle MODCOD doit permettre de déterminer le MODCOD de la prochaine BBFRAME. Étant donné que la définition précédente des règles utilisateurs ne prend pas en compte le débit atteignable, ce rôle doit être dévolu à la règle MODCOD. Si P_m est la taille de la charge utile d'une BBFRAME de MODCOD m et τ_m son temps de transmission, on définit le débit utile comme :

$$R_m = \frac{P_m}{\tau_m}$$

Ce débit est indépendant du temps, étant associé à une file de l'ordonnanceur, contrairement au débit atteignable dans les règles définies dans les systèmes terrestres. Il joue cependant un rôle similaire, puisqu'il exprime la capacité d'un utilisateur à écouler son trafic plus ou moins rapidement. Ici, la variabilité s'observera plutôt sur le débit entrant dans chaque file de QOS, qui évoluera avec les conditions météorologiques des Terminaux, en plus de l'évolution de leur trafic :

Forts de cette constatation, nous formulons la règle MODCOD de la manière suivante :

$$m^* = \arg \max_m R_m \sum_{k \in \nu_m} \delta_k(n) c_k(n) \quad (3.3)$$

Cette règle prend donc en compte le MODCOD, mais s'exprime également en fonction des règles utilisateur exprimées plus haut. Cette réutilisation permet une cohérence des décisions entre le choix du MODCOD et des paquets. Notons que l'implantation d'une telle règle est très simple, puisqu'il suffit d'un calcul par MODCOD pour déterminer le MODCOD de la prochaine BBFRAME.

Dans le cas de PF, nous obtenons la règle MODCOD suivante :

$$m^* = \arg \max_m R_m \sum_{k \in \nu_m} \frac{\delta_k(n)}{\bar{r}_k(n)} \quad (3.4)$$

La formulation que nous avons adoptée ne prend pas en compte le déclassement, puisque seules les files du MODCOD m sont considérées. Ce choix découle d'un constat pragmatique : si le déclassement peut être utilisé à faible charge, il est relativement peu présent à forte (> 0.8) charge, comme indiqué dans [6] [28]. Or, les mécanismes

d'ordonnancement ne montrent leur potentiel qu'à forte charge, là où les décisions d'ordonnancement ont un impact important sur le système. Nous avons donc considéré le reclassement comme potentiellement présent de manière marginale, mais l'avons négligé dans le calcul.

3.3 Règles d'ordonnancement adaptées au satellite

En suivant les principes d'adaptation présentés précédemment, nous montrons ici comment M-LWDF et EXP-PF ont été adaptés à notre contexte, tout en restant cohérents avec la définition des règles retenues pour les systèmes terrestres. Cette cohérence est nécessaire afin d'assurer la crédibilité des règles d'ordonnancement.

3.3.1 M-LWDF et EXP-PF : formulation

M-LWDF utilise le temps d'attente des paquets en tête de file d'attente comme métrique principale, combinée avec un paramètre de QoS ainsi que le débit instantané. Son expression est donnée par l'équation (2.4), et se compose de trois termes :

- le temps d'attente moyen en tête de file $\bar{w}_k(n)$;
- le débit atteignable $r_k(n)$;
- un paramètre de QoS, γ_k .

Une des particularités de la formulation de [27] est de proposer la valeur suivante pour γ_k :

$$\gamma_k = \frac{a_k}{\bar{r}_k(n)}$$

Où a_k est une priorité de QoS. Cette formulation permet d'introduire l'expression de PF dans M-LWDF, qui sera ainsi également adapté aux flux élastiques. Pour ce qui est de la priorité de QoS, les auteurs de [27] proposent de l'adapter aux contraintes de débit définies pour l'utilisateur considéré. Supposons que la proportion maximale de paquets dont le délai dépasse un seuil fixé, noté $w_{k,max}$, soit $p_{k,max}$. On peut alors écrire la relation suivante :

$$P[w_k(n) > w_{k,max}] = p_{k,max}$$

La valeur de a_k peut être alors définie comme :

$$a_k = \frac{-\ln(p_{k,max})}{w_{k,max}}$$

Étant donné que γ_k dépend, selon la définition donnée, du temps, nous noterons à présent $\gamma_k(n)$ cette valeur. Suivant l'approche décrite dans le chapitre précédent, nous proposons

la règle utilisateur suivante pour M-LWDF :

$$c_k(n) = \gamma_k(n)\bar{w}_k(n) \quad (3.5)$$

Une fois de plus, le débit atteignable sera porté par la règle MODCOD, dont l'expression sera :

$$m^* = \arg \max_m R_m \sum_{k \in \nu_m} \delta_k(n)\gamma_k(n)\bar{w}_k(n)$$

On peut remarquer que le fait de prendre en compte le délai en tête de file suppose que la décision d'ordonnancement ne prendra pas en compte le délai des paquets plus loin dans la file. Or, puisqu'une BBFRAME dispose d'une charge utile importante, ces paquets pourront éventuellement être encapsulés dans cette même BBFRAME. Il sera donc intéressant d'observer les différences de délai entre les paquets au sein d'une BBFRAME.

La règle EXP-PF, définie dans [47], permet de réduire l'influence du délai en tête de file dans l'expression de la règle, et ainsi de prendre en charge de façon plus efficace des trafics ayant des caractéristiques disparates, comme dans notre cas d'étude. L'idée de cette règle est d'appliquer une valorisation exponentielle à la différence entre le délai mesuré et le délai moyen (toujours en tête de file). L'expression donnée dans [47] est la suivante :

$$i^* = \arg \max_i \frac{r_i(n)}{\bar{r}_i(n)} \exp\left(\frac{a_i\bar{w}_i(n) - \overline{aW}}{1 + \sqrt{\overline{aW}}}\right)$$

Avec :

$$\overline{aW} = \frac{1}{N_{rt}} \sum_{i=0}^{N_{rt}} a_i\bar{w}_i(n)$$

Cette somme ne se fait que sur les N_{rt} flux temps réel, par définition. Les autres valeurs sont les mêmes que précédemment. De façon très similaire à M-LWDF, nous proposons la règle utilisateur suivante :

$$c_k = \frac{1}{\bar{r}_k(n)} \exp\left(\frac{a_k\bar{w}_k(n) - \overline{aW}}{1 + \sqrt{\overline{aW}}}\right)$$

La règle MODCOD pour notre adaptation de EXP-PF suivra l'équation (3.3), tout comme M-LWDF. Le terme \overline{aW} sera la somme sur les files EF et AF actives, c'est-à-dire où $\delta_k(n) = 1$.

Ces trois nouvelles règles, nous permettent de définir le cadre des règles d'ordonnancement dans notre contexte. Cependant, elles ne constituent qu'une partie de l'algorithme en lui-même, qu'il reste donc à spécifier, notamment concernant le choix des paquets

encapsulés dans la BBFRAME, en appliquant la règle utilisateur.

3.3.2 Algorithme utilisant les règles d'ordonnement

Une fois les règles utilisateur et MODCOD définies, il nous faut définir un algorithme d'ordonnement adapté. Afin de refléter la dépendance entre choix des paquets et choix de la BBFRAME, nous avons utilisé la règle utilisateur dans la règle MODCOD. Ainsi, cette dernière permet de prendre en compte à la fois les contraintes de QoS et l'efficacité spectrale. Cependant, une fois le choix du MODCOD effectué, il faut encore utiliser les règles utilisateur pour sélectionner les paquets qui seront transmis. Ce problème est assez différent de l'utilisation classique des règles d'ordonnement, puisqu'un certain nombre de paquets provenant de la *même file* peuvent être encapsulés dans une même BBFRAME.

Nous proposons donc l'algorithme suivant. A chaque instant d'ordonnement, toutes les règles utilisateur et MODCOD sont évaluées. Comme spécifié plus haut, le nombre de MODCOD utilisés simultanément est faible, le nombre d'opérations sera donc de quelques dizaines par BBFRAME, ce qui semble tout à fait raisonnable pour un ordonnement en temps réel. Une fois ces calculs effectués, la prochaine BBFRAME est déterminée en calculant toutes les règles MODCOD et en choisissant la meilleure, selon la formule (3.3). Les règles utilisateurs pour ce MODCOD sont alors triées par ordre décroissant, puis on vide la file ayant la meilleure règle utilisateur dans la BBFRAME. Si la BBFRAME n'est pas remplie, on procède de la même manière avec les autres files utilisateur jusqu'à ce que la BBFRAME soit pleine, ou que toutes les files utilisateur de ce MODCOD soient vides. Dans ce dernier cas, la même procédure de choix du MODCOD et des paquets est répétée, mais uniquement avec les files ayant un meilleur MODCOD, afin que leurs paquets puissent être déclassés, et avec un volume de BBFRAME égal au volume restant.

Cette méthode comporte plusieurs avantages. Tout d'abord, elle permet d'effectuer un ordonnement BBFRAME et paquets cohérent, car mené selon les mêmes critères (les règles utilisateur). La complexité totale de cet algorithme, sans compter le déclassement est de $o(N + M + Q \ln(Q))$, dans le cas où un algorithme de tri rapide est utilisé pour les règles utilisateurs. L'un des avantages principal de cet algorithme est toutefois de simplifier la fragmentation. En effet, il est rare que le volume total des paquets encapsulés égale la charge utile de la BBFRAME. GSE permet de fragmenter les paquets IP, mais cette opération introduit de l'overhead. Aussi, minimiser la fragmentation constitue-t-il une priorité afin d'éviter une baisse de l'efficacité spectrale ; notre algorithme n'occasionne, au maximum, qu'un seul fragment par BBFRAME, pour un coût calculatoire faible, ce qui semble être un bon compromis.

Nous pouvons cependant remarquer que cette règle pourra être inéquitable entre

les QoS, puisqu'elle privilégie la file, et non les paquets, ayant la meilleure règle. Par rapport à un RR, dont l'ordre serait déterminé par les règles utilisateur, cette solution aura également tendance à augmenter les différences de délai entre les paquets au sein d'une BBFRAME. Il faudra donc être attentif à ce problème.

3.4 Ordonnancement et fonctions d'utilité

Les fonctions d'utilité constituent une autre solution d'ordonnancement pertinente pour notre contexte. Si elles sont bien plus complexes à mettre en place que les règles d'ordonnancement, leur flexibilisé et l'apport de la formulation mathématique constituent un avantage indéniable.

La différence majeure avec les règles d'ordonnancement est la capacité de la formulation sous forme de problème d'optimisation à *anticiper* l'impact des décisions d'ordonnancement, de manière à mener le système vers un état optimal défini à l'avance. Les fonctions d'utilité offrent la possibilité d'explicitier le compromis entre QoS et efficacité spectrale, grâce à une formulation prenant en compte conjointement de nombreux paramètres dans une seule fonction.

Nous présentons dans cette partie la formulation de notre problème d'optimisation, reposant sur des fonctions d'utilité, puis l'implantation que nous avons choisi, en insistant sur l'impact de la méthode de résolution choisie sur les possibilités d'implantation.

3.4.1 Formulation du problème

Nous définissons tout d'abord une métrique moyenne générique, \bar{x}_k , d'intérêt pour le système. Cette métrique pourra représenter, par exemple, la conformité de l'ordonnancement aux garanties de QoS (débit garanti, délai garanti). Cette métrique est une valeur moyenne sur un temps considéré long devant celui de l'ordonnancement. Même si l'optimisation conjointe de plusieurs métriques est éventuellement possible, comme indiqué dans [55], cette technique se révèle plus complexe, et [55] montre qu'il est préférable de ramener le problème à une unique métrique. En effet, lorsqu'une optimisation est conduite sur deux variables en même temps, par exemple débit et délai, l'un ne doit pas prendre le pas de manière systématique sur l'autre. De plus, l'ajout de variables, et donc de directions à l'optimisation du service reçu par une file rend l'optimisation bien plus complexe, et les bénéfices sont difficiles à estimer [55]. Nous limiterons donc notre analyse à une seule métrique identique pour toutes les files. Cette solution réduit certes la flexibilité, mais permet de limiter la complexité de l'algorithme.

Posons u_k la fonction d'utilité associée à l'utilisateur k , et définissons :

$$X = (\bar{x}_0, \dots, \bar{x}_{N-1})$$

le vecteur des métriques moyennes, et $U(X)$ l'utilité du système associée à ces valeurs. Afin de définir l'optimal, nous ne considérons pas, dans un premier temps, l'aspect temporel. Cette démarche peut sembler en contradiction avec ce qui a déjà été dit jusqu'ici, mais elle est nécessaire afin de définir la direction de la solution optimale. C'est vers cet objectif que va tendre l'ordonnancement. En effet, nous pouvons exprimer l'objectif à *long terme* de l'ordonnancement comme la solution du problème suivant :

$$\begin{aligned} & \underset{X}{\text{maximiser}} & U(X) &= \sum_{k=1}^{N-1} u_k(\bar{x}_k) \\ & \text{avec} & & X \in C \end{aligned} \tag{3.6}$$

Comme nous avons remarqué dans le chapitre 2, la solution de ce problème peut être trouvée à l'aide de méthodes Lagrangiennes [76], mais cette solution n'est d'aucune aide, étant donné le caractère à la fois limité et changeant de la capacité du système, C , sans oublier que la charge de chaque MODCOD évolue également au cours du temps. Le problème (3.6) définit donc une direction vers laquelle il faudra déplacer le système. Aussi notre problème se résume-t-il à trouver un algorithme permettant, à chaque instant d'ordonnancement, de se rapprocher le plus possible de cet optimum. Cette technique fait correspondre un algorithme de résolution avec le comportement du système.

Soit deux états successifs X et X' , séparés d'un nombre quelconque d'instant d'ordonnancement, notre méthode correspondra à la solution de l'équation suivante :

$$\max_{X'} U(X') - U(X) \tag{3.7}$$

Sachant que la différence entre X et X' sera bornée par la vitesse d'évolution du système. La solution de (3.7) donne la direction de l'ordonnancement entre les instants correspondant aux états X et X' .

3.4.2 Méthodes de résolution

Nous développons ici deux méthodes, reposant chacune sur une approximation différente permettant de faire correspondre le comportement de notre système avec la résolution du problème d'optimisation défini en (3.6). Notre méthode est guidée par l'observation suivante. Étant donné que le système ne peut être déplacé instantanément

à l'optimum, il doit constamment se déplacer vers cet optimum, et ce le plus rapidement possible. Comme nous l'avons vu, la direction de cet optimum change, il sera donc nécessaire de recalculer régulièrement sa direction, de manière à éviter un ordonnancement allant dans la mauvaise direction. Autrement dit, le système se déplacera dans la direction de l'optimum tant que cette direction ne change pas significativement. Toute la difficulté de notre problème réside dans l'estimation de cette durée, que nous noterons T_s , la période d'ordonnancement. Cette durée peut être exprimée, conformément à notre modèle de mesure du temps, de la façon suivante :

$$T_s = \sum_{n=0}^{N_{T_s}-1} \tau(n) \quad (3.8)$$

Où $\tau(n)$ est la durée de la BBFRAME émise à l'instant n , et N_{T_s} est le nombre de BBFRAME considérées. La période d'ordonnancement, définie ainsi, n'est pas nécessairement strictement constante : si N_{T_s} est fixé, T_s sera variable. On peut également définir une valeur fixe de T_s , où le nombre – potentiellement non entier – de BBFRAME N_{T_s} par période d'ordonnancement sera variable.

BBFrame-by-BBFrame Scheduling

Dans cette approche, nous présentons une extension les travaux présentés dans [26] [58] et [28]. Cette méthode, dite de *poursuite*, recalcule la direction de l'optimum à chaque instant d'ordonnancement, et se présente ainsi comme celle permettant de suivre au plus proche l'optimum. Du fait de sa précision, cette méthode suppose un coût de calcul plus élevé, puisqu'il faudra résoudre un problème de la forme de (3.7) à chaque instant d'ordonnancement. Formellement, cette technique correspond à prendre $N_{T_s} = 1$, et résoudre ainsi, à chaque BBFRAME, le problème suivant :

$$\max_{X(n+1)} U(X(n+1)) - U(X(n)) \quad (3.9)$$

Cette solution sera dénommée BBFrame-by-BBFrame Scheduling (BBS) dans la suite de cette étude.

Periodic Scheduling

Dans cette seconde approche, nous relaxons la contrainte sur la durée de T_s , en supposant qu'une durée équivalente à quelques BBFRAMES ne modifie pas significativement la précision de nos calculs. Plus spécifiquement, on considère qu'il est possible de prédire

avec suffisamment de précision l'état du système au bout de T_s , et d'estimer ainsi le gain d'utilité possible. Cette solution sera appelée BBFrame Periodic Scheduling (BPS) dans la suite de cette étude. La direction de l'objectif est alors donnée par l'expression suivante, où $N_{T_s} > 1$:

$$\max_{X(n+N_{T_s})} U(X(n + N_{T_s})) - U(X(n)) \quad (3.10)$$

Si l'ordonnancement est déterminé pour les T_s prochaines BBFRAME, la charge de calcul peut être mieux répartie que dans le cas de BBS. De plus, avoir une tâche périodique, déterminant les actions du système pour une période déterminée est un avantage certain pour l'implantation.

3.4.3 Algorithme reposant sur le délai

BBS et BPS permettent de construire un algorithme d'ordonnancement sous-optimal, utilisant un large panel de métriques et de fonctions d'utilité. Nous présentons ici une application de ces algorithmes à notre contexte d'ordonnancement, mais cette méthode est suffisamment générique pour être réutilisée dans un autre contexte (QoS, définition des utilisateurs).

Ainsi, nous avons choisi le temps d'attente moyen d'un paquet dans une file comme métrique de référence. Cette métrique s'accorde bien avec la définition DIFFSERV de QoS que nous avons utilisée : chaque PHB peut être défini en fonction de ce temps d'attente moyen. Nous donnons cependant l'esquisse d'une solution utilisant le débit moyen, afin de prouver la flexibilité de notre approche.

Estimation du temps d'attente

Le temps d'attente à la fin de la période d'ordonnancement est donné par la Loi de Little :

$$w_k(n + N_{T_s}) = \frac{l_k(n) + T_s \lambda_k(n) - s_k}{\bar{r}_k(n)} \quad (3.11)$$

Où sont définis, pour l'utilisateur k :

- $l_k(n)$ la taille de la file d'attente à l'instant d'ordonnancement n ;
- $\lambda_k(n)$ est le débit moyen d'entrée ;
- $\bar{r}_k(n)$ est le débit moyen de sortie ;
- s_k est le volume de données servi pendant T_s .

Les débits moyens sont mesurés à l'aide d'une moyenne exponentielle, comme indiqué dans le Chapitre 2.

En utilisant la formule (2.1), nous obtenons l'expression suivante :

$$\bar{w}_k(n + N_{T_s}) = \alpha(n)\bar{w}_k(n) + (1 - \alpha(n))w_k(n + N_{T_s}) \quad (3.12)$$

En supposant que la charge globale du système est stable, et en supposant un système sans pertes de paquets, nous pouvons utiliser $\bar{r}_k(n)$ comme estimateur de $\lambda_k(n)$. La précision de cette évaluation sera fonction de la période T_s .

Le temps d'attente du système w est défini par :

$$w = (\bar{w}_1, \dots, \bar{w}_N)$$

L'utilité associée est :

$$U(w) = \sum_k u_k(\bar{w}_k)$$

BBFrame-by-BBFrame Scheduling

Dans BBS, nous avons $N_{T_s} = 1$, c'est-à-dire que T_s vaut la durée d'une BBFRAME. Du point de vue du système, cette durée est faible, et le calcul de (3.9) peut se simplifier grâce à l'approximation du premier ordre suivante :

$$U(w(n+1)) \approx U(w(n)) + U'(w(n))dw$$

La direction de l'optimum est déterminée en maximisant le dernier terme, appelé *utilité marginale*. Dans cette approximation, on suppose que le changement introduit par le déplacement du système de $w(n)$ vers $w(n+1)$ est suffisamment faible, ce qui est raisonnable étant donné le volume et le temps d'émission d'une BBFRAME. On retrouve ici l'algorithme du gradient, décrit dans [76], une BBFRAME correspondant à un pas de l'algorithme, où nous suivons la direction de l'optimum à chaque itération.

L'expression de $dw = (d\bar{w}_1, \dots, d\bar{w}_N)$ peut être déterminée en dérivant (3.12) selon s , puisque le temps d'attente estimé est une fonction du volume servi :

$$d\bar{w} = -\frac{(1 - \alpha^{\tau(n)})}{\lambda_k(n)} ds_k$$

Où ds_k est un petit volume servi, correspondant à la différenciation de \bar{w}_k .

Nous pouvons exprimer l'utilité marginale totale, pour le MODCOD m comme il suit :

$$\sum_{k \in \nu_m} -u'_k(\bar{w}_k(n)) \frac{(1 - \alpha^{\tau(n)})}{\lambda_k(n)} ds_k \quad (3.13)$$

Maximiser cette expression en fonction de $ds = (ds_1, \dots, ds_N)$ et m donne la direction de l'optimum. Le nombre de MODCOD simultanément utilisés dans un système étant faible, il est possible de réaliser ce calcul pour tous les MODCOD, de manière exhaustive. Il reste donc le calcul de l'utilité marginale maximale pour le MODCOD m , problème que l'on peut exprimer de la manière suivante :

$$\begin{aligned} \max_{ds} \quad & \sum_{k \in \nu_m} A_k ds_k \\ \text{avec} \quad & \sum_{k \in \nu_m} ds_k \leq P_m \end{aligned} \quad (3.14)$$

Où P_m est la charge utile d'une BBFRAME de MODCOD m , et A_k est une constante :

$$A_k = -u'_k(\bar{w}_k(n)) \frac{(1 - \alpha^{\tau(n)})}{\lambda_k(n)}$$

En considérant s comme un volume continu, et non un ensemble de paquets, il est possible de résoudre de manière simple ce problème, en appliquant un algorithme très similaire à celui décrit pour les règles d'ordonnancement, A_k jouant ici le rôle de la règle utilisateur. Cette solution apporte les mêmes bénéfices en termes de passage entre la solution continue, s , et la solution discrète, composée de paquets.

Il est intéressant de noter qu'en sélectionnant $u'_k(\bar{w}_k) = -\bar{w}_k$ nous obtenons une expression très proche de M-LWDF, où la somme de (3.13) devient :

$$\sum_{k \in \nu_m} \bar{w}_k(n) \frac{(1 - \alpha^{\tau(n)})}{\lambda_k(n)} ds_k$$

L'utilisation du débit moyen, plus simple à estimer, mène à l'expression suivante :

$$\sum_{k \in \nu_m} u'_k(\bar{r}_k) dr_k(n)$$

Où dr_k est le débit instantané correspondant à un volume ds_k . Un choix d'une fonction d'utilité tel que $u(\bar{r}) = \log(\bar{r})$ permet ici de retrouver PF, montrant la flexibilité et le caractère générique de notre formulation.

BBFrame Periodic Scheduling

Dans BPS, la durée T_s est allongée, il n'est donc plus possible de faire l'approximation au premier ordre de BBS. La direction de l'optimum est donc :

$$\begin{aligned} & \underset{S}{\text{maximiser}} && \sum_m \sum_{k \in \nu_m} u_k(\bar{w}_k(n + N_{T_s})) \\ & \text{avec} && \sum_m \sum_{k \in \nu_m} \frac{s_k}{R_m} = T_s \end{aligned} \quad (3.15)$$

Où S est le vecteur des volumes servis s_k pendant T_s . Résoudre ce problème, qui appartient à la programmation convexe non linéaire séparable, est une tâche complexe. Étant donnée la convexité de l'ensemble de définition de S , et en supposant u concave, on peut montrer qu'il existe une unique solution [76], calculable à l'aide de méthodes Lagrangiennes. Un algorithme est proposé dans [78], reposant sur la détermination itérative des ensembles des contraintes saturées. Les résultats quant à la complexité, exposés dans [78] montrent des résultats satisfaisants, compatibles avec un fonctionnement en temps réel.

La résolution de ce problème ne donne pas de solution d'ordonnancement, mais un volume par utilisateur, s_k^* , qu'il faudra servir pendant T_s , ou bien pendant les N_{T_s} prochaines BBFRAME.

C'est ici qu'apparaît un problème majeur de l'ordonnancement tel que nous l'avons défini : notre solution optimale, S^* est exprimée en nombre réel. Un premier arrondi à l'entier le plus proche permet d'obtenir un nombre entier d'octets (ou bits), qui, étant donné les tailles classiques des paquets, ne devrait pas biaiser fortement notre algorithme. Mais un deuxième problème surgit alors : comment servir exactement s_k^* octets en T_s , sachant que notre système ne permet d'émettre que des BBFRAME dont la charge utile est fixée à l'avance. Qui plus est, comment minimiser le nombre de fragments de paquets ? S'il est possible, avec GSE, de servir exactement un volume s_k^* dans chaque file, cette solution mènera à de nombreux fragments, qui auraient pu être évités. En résumé, le passage de S^* à une véritable solution d'ordonnancement est très complexe, d'autant plus que la solution d'ordonnancement doit être déterminée à chaque BBFRAME en fonction du volume déterminé par S^* résiduel.

Afin d'éviter de complexes calculs combinatoires, et pour permettre un *étalage* des BBFRAME sur la période T_s , nous proposons une adaptation du DRR à notre contexte, reliant la solution optimale S^* aux paramétrage du DRR, et plus exactement du quantum. Cet algorithme est construit avec le principe suivant : on associe à chaque utilisateur un déficit d_k , ainsi qu'un quantum ϕ_k , ajouté à chaque tour du RR. On définit un RR entre les MODCOD, ainsi qu'un RR entre les QOS. A l'instant d'ordonnancement, la solution

d'ordonnancement est déterminée selon le principe d'un DRR : le MODCOD suivant, dans l'ordre du RR est sélectionné, et son déficit D_m , c'est-à-dire la somme des déficits de ses utilisateurs, est déterminé :

$$D_m = \sum_k d_k$$

La somme porte ici uniquement sur les utilisateurs du MODCOD m , soit Q files. Si ce déficit cumulé dépasse ou égale la capacité d'une BBFRAME, alors cette BBFRAME est choisie pour la prochaine solution d'ordonnancement, soit la condition suivante :

$$D_m \geq C_m$$

Les paquets sont ensuite servis selon un DRR, tant que la BBFRAME n'est pas pleine. Cette étape concerne donc plusieurs paquets par file, contrairement au DRR original. Notons que cette solution pourrait éventuellement mener à un déficit négatif. Une file ayant un déficit négatif à l'instant d'ordonnancement reçoit toujours son quantum, mais n'est pas considérée dans la solution d'ordonnancement. La relation entre quantum et volume s_k^* est la suivante :

$$\forall k, k' \frac{\phi_k}{s_k^*} = \frac{\phi_{k'}}{s_{k'}^*}$$

De cette manière, la détermination d'un seul ϕ_k permet le calcul de tous les autres. Cet algorithme est représenté sur la figure 3.2.

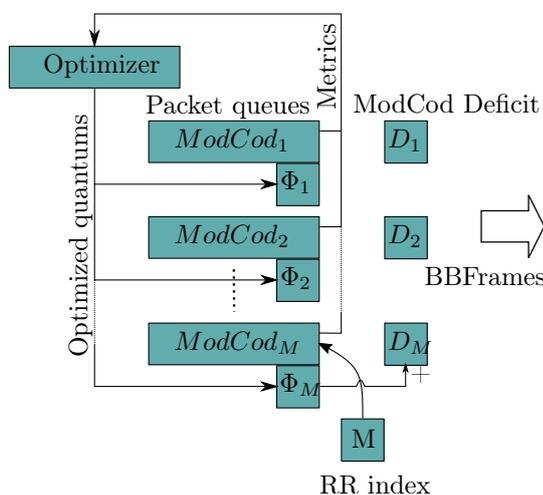


FIGURE 3.2 DRR adapté aux BBFRAME

Cette solution possède un avantage important dans la régularité de son calcul : tous les T_s , un calcul d'optimisation est effectué, dont la complexité de l'algorithme de résolution

est connu. Le DRR proposé possède une complexité très faible, en temps constant. S'il ne possède pas d'avantage évident sur BBS, BPS est plus simple à implanter, tout en remplissant les critères définis au début de cette étude (prise en compte du caractère dynamique de la capacité et de la demande à travers T_s). Il faudra cependant prêter attention aux problèmes de passage continu-discret, ainsi qu'à la différence entre volume optimal et volume effectivement servi à chaque BBFRAME, afin de garantir la pertinence de notre approche.

3.5 Évaluation de performances

Nous présentons dans cette partie une évaluation des performances des algorithmes d'ordonnancement sur la voie Aller présentés dans cette étude. Notre objectif est double : montrer la faisabilité des solutions proposées dans un contexte satellite, à travers une étude des performances, de la complexité, et de capacité des algorithmes à s'adapter au scénario, et déterminer les solutions qui semblent le mieux répondre aux objectifs définis.

Les solutions présentées ont toutes été implantées dans le simulateur NDES [79], une bibliothèque de simulations de files d'attente, en C, développée au laboratoire IRIT.

3.5.1 Scénarios

Nous définissons ici les contraintes de QoS correspondant à notre modèle, le délai maximum $w_{k,max}$ d'un paquet dans la file, et la gigue δ_k . Ces paramètres, ainsi que la répartition du trafic dans chaque classe, sont donnés dans la table 3.1. Notons que, si ces paramètres diffèrent de ceux présentés dans [80] pour EF, rien n'empêche l'adaptation de notre solution à ces paramètres. Cette répartition correspond à celle d'un service de type «triple-play» qui est le coeur de marché de notre système. C'est donc le scénario nominal dans lequel il doit être utilisé.

Classe	EF	AF	BE
charge	0.1	0.3	0.6
w_{max} (s)	0.05	0.1	0.2
δ_{max} (s)	0.01	0.05	0.1

TABLE 3.1 Paramètres de QoS

Nous définissons également, à des fins d'analyse, deux scénarios de répartition de QoS :

- Un scénario «iso QoS» où chaque classe de QoS contribue de manière égale à la charge. Ce scénario permet d'observer le comportement quant à l'efficacité spectrale de l'algorithme, indépendamment de la qualité de service.
- Un scénario où l'application majoritaire n'est plus l'accès IP mais la téléphonie IP. La charge engendrée par la classe EF sera donc ici majoritaire. Ce scénario peut correspondre à un système de type Inmarsat, où la majorité des utilisateurs utiliseraient le service de voix. Bien que ce scénario soit hors du cadre système que nous nous sommes défini, il permet d'analyser le comportement des algorithmes d'ordonnancement dans un cas « limite ».

On notera que les deux derniers scénarios ne correspondent pas à un scénario réel, mais permettent d'analyser plus finement le comportement des algorithmes.

Ainsi, nous attacherons-nous à observer le délai moyen dans les différentes files d'attente, afin de connaître la compatibilité de notre algorithme avec ces contraintes.

Afin de tester nos algorithmes dans des scénarios différents, nous proposons trois répartitions pour les MODCOD, l'une correspondant au cas de référence, ou de ciel clair, à la table 3.2, l'autre correspondant à un cas très dégradé, ou de forte pluie, dans la table 3.3. Une dernière répartition, dite «iso-MODCOD» nous permettra d'observer le comportement de l'algorithme quant au respect de la Qualité de service, indépendamment de l'efficacité spectrale. Si seul le premier scénario correspond à un cas plausible, le but des deux autres répartitions est d'observer l'influence de la répartition des MODCOD sur le système. Nous pouvons noter que seuls quatre MODCOD sont utilisés. Ce nombre relativement faible relève d'une contrainte d'implantation : si le nombre de MODCOD potentiel de DVB-S2 est relativement élevé, le nombre de MODCOD simultanément utilisés est généralement faible.

#	ModCod	charge
MC0	QPSK 1/3	0.01
MC1	8-PSK 3/5	0.19
MC2	8-PSK 5/6	0.3
MC3	16-APSK 2/3	0.5

TABLE 3.2 distribution *Ciel clair*

La charge globale du système est fixée à 0.9 : en considérant la loi d'arrivée des paquets pour chaque QoS et MODCOD, il est aisé de calculer le débit entrant de chaque file d'attente. Cette charge élevée, où le système est proche de la saturation, permet d'observer de façon plus nette les différences de performances entre les algorithmes. Les

#	ModCod	charge
MC0	QPSK 1/3	0.5
MC1	8-PSK 3/5	0.3
MC2	8-PSK 5/6	0.19
MC3	16-APSK 2/3	0.01

TABLE 3.3 distribution *Forte pluie*

ressources allouées au système étant particulièrement coûteuses, il est nécessaire d'en faire un usage aussi efficace que possible. Aussi, en ce qui concerne l'ordonnancement, on souhaite utiliser le système à une charge proche de la saturation, afin de maximiser le revenu engendré par les ressources fréquentielles allouées, d'où une charge élevée. Le débit sortant étant donné par le MODCOD, le calcul de la charge pour chaque file, et par conséquent de la charge globale, peut être fait simplement. Le temps de simulation est de 100 secondes, ce qui semble suffisant pour pouvoir ignorer les effets de la période transitoire, en tenant compte des propriétés du trafic. La valeur de T_s choisie pour BPS est de 32 ms, ce qui correspond à quelques dizaines de BBFRAME. Les arrivées de paquets suivent une loi de Poisson par file, étant donné qu'elles correspondent à un multiplexage de très nombreux flux, les arrivées peuvent être considérées comme indépendantes.

Afin de répondre aux contraintes de QOS, nous proposons la fonction d'utilité suivante, paramétrée par le délai :

$$u_k(\bar{w}_k) = -\frac{1}{\delta_{k,max}^2} (w_{k,max}/2 - \bar{w}_k)^2 \quad (3.16)$$

Cette famille de fonctions a l'avantage de ne pas nécessiter de paramétrage autre que celui donné par la QOS (par exemple l'équité, comme dans [7]). Leur comportement, présenté à la figure 3.3 diffère des fonctions d'utilité généralement utilisées, puisque, en-deçà d'un certain délai, il est moins intéressant (ou utile) de servir le trafic temps réel par rapport au trafic BE. Ainsi, ces fonctions permettent, sur un système peu chargé, de répartir la capacité de façon à garantir un délai constant aux flux temps réel, tout en répartissant mieux la capacité qu'une priorité fixe. Cette propriété permet une meilleure cohabitation de trafic ayant des besoins en QOS différents.

Les algorithmes présentés ici sont, dans un premier temps :

- les algorithmes à critère explicite [4] et [5] ;
- les règles d'ordonnancement définies M-LWDF et EXP-PF ;
- BBS et BPS, comme définis plus haut, avec le délai comme métrique.

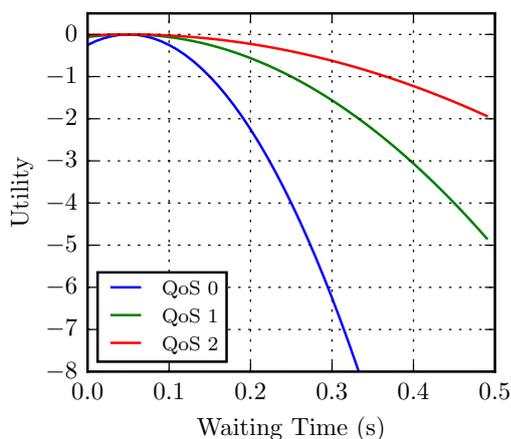


FIGURE 3.3 Fonctions d'utilité valorisant le temps d'attente

Nous indiquons, lorsqu'elles sont pertinentes, les mesures de complexité et discutons le passage continu-discret pour BPS.

3.5.2 Scénario ciel clair

Cette première évaluation a pour but de déterminer si les algorithmes présentés permettent d'ordonnancer le trafic entrant, dans un cas de référence, tout en respectant les contraintes de QoS. Les résultats de simulations sont présentés sur la figure 3.4. Le délai moyen est représenté, en échelle logarithmique, en fonction du MODCOD et de la QoS. Les contraintes de QoS sont représentées en traits rouges.

La figure 3.4 présente trois types d'algorithmes. Tout d'abord, les solutions WRR [4] et UBMT [5] (3.4a 3.4b) favorisent de façon claire les meilleurs MODCOD au détriment des plus faibles, qui sont aussi les moins chargés. Si UBMT ne dépasse que peu les contraintes de QoS pour le MODCOD le moins chargé, WRR ne permet de les satisfaire pour le BE pour quasiment aucun MODCOD¹. Ce comportement inéquitable peut créer une situation instable, comme sur la figure 3.4b, où le délai de la QoS BE augmente continuellement pendant la simulation. Le caractère instable de cette solution peut également être observé dans les intervalles de confiances du WRR, particulièrement importants. Cette situation montre à quel point il est complexe, pour un algorithme, de s'adapter à un scénario en particulier, et plus spécifiquement à de grandes différences de débit entrant entre MODCOD. Les suppositions de [4] sur la nature du trafic ne sont tout simplement pas vérifiées ici, grevant le fonctionnement de l'algorithme d'ordonnancement. Même s'il est

1. Le paramètre d'équité a été pris égal à 1

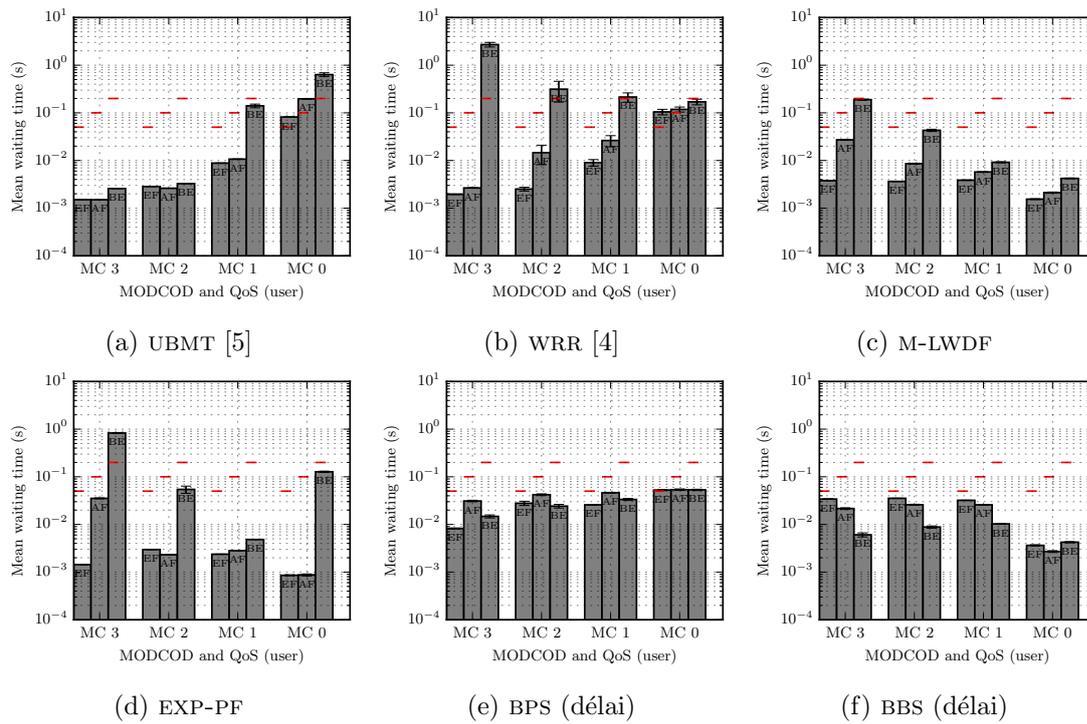


FIGURE 3.4 Temps d'attente moyen pour plusieurs algorithmes de référence

possible d'adapter l'équité au scénario, cette adaptation se fera de manière *ad hoc* et n'aura donc aucune valeur dans un autre scénario. Nous pouvons noter qu'UBMT ne comporte pas ce genre d'adaptation, et permet un fonctionnement tout à fait correct, où seul le MODCOD le moins chargé ne voit pas ses critères de QOS respectés. Nous pouvons également noter le peu de différence entre le traitement de EF, et BE, ainsi que les intervalles de confiance faibles.

Les règles d'ordonnancement M-LWDF et EXP-PF (3.4c 3.4d) offrent une meilleure performance, même si l'on peut noter une différence importante de délai au sein d'une même QOS. On peut remarquer les résultats décevants de EXP-PF au regard des attentes de cette règle, à savoir une meilleure gestion des trafics hétérogènes : le délai pour le trafic BE du MODCOD 3, qui représente la part la plus importante du trafic total, est hors de ses contraintes de QOS. Cette règle ne privilégie pas suffisamment cette QOS au profit des autres, qui voient leur performance augmenter par rapport à M-LWDF. Cette dernière règle offre une performance compatible avec la QOS définie, ce qui en fait un candidat intéressant pour notre système. La simplicité de calcul de M-LWDF, est un avantage supplémentaire, étant donné ses performances comparativement à celles d'EXP-PF.

Le troisième type d'algorithme, BBS et BPS (3.4e 3.4f), semble offrir de bonnes performances, comparables voire meilleures que M-LWDF et UBMT. On peut remarquer que ces algorithmes ont une performance mieux répartie entre les MODCOD et les QOS, ne favorisant aucun en particulier. Cette propriété est particulièrement intéressante, sachant que ces systèmes seront amenés à fonctionner dans des conditions potentiellement très hétérogènes. On peut noter que BPS est légèrement moins performant que BBS, dont le fonctionnement est bien plus complexe, offrant un aperçu du compromis entre performance et complexité.

Ainsi, dans ce cas de référence, nous retenons les solutions qui semblent convenir aux critères de performance que nous nous sommes fixés : UBMT, M-LWDF et BPS. Le WRR de [4] ne permet pas d'obtenir des résultats satisfaisants, tandis que EXP-PF ne présente par d'intérêt majeur par rapport à M-LWDF. Le cas de UBMT est plus complexe, et aurait demandé une analyse plus poussée.

3.5.3 Scénario forte pluie

Dans ce scénario, nous examinons comment les algorithmes retenus se comportent face à un cas limite, où la répartition de charge des MODCOD est inversée par rapport au cas précédent. Les résultats de ces simulations sont présentés sur la figure 3.5.

La particularité de ce scénario réside dans la répartition de la charge : le MODCOD le plus faible est celui devant faire face à la charge la plus importante (50% du trafic).

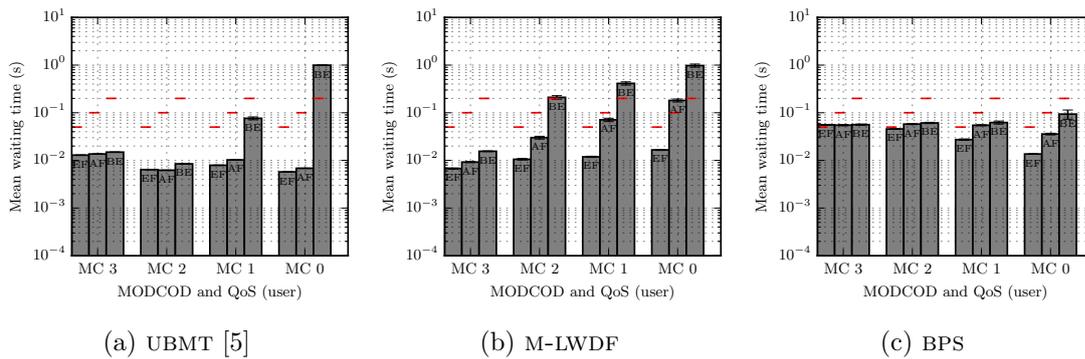


FIGURE 3.5 Temps d'attente moyen, scénario forte pluie

Ainsi, dans ce scénario, les algorithmes privilégiant l'efficacité spectrale risquent de ne pas pouvoir ordonnancer efficacement ce dernier MODCOD.

De fait, le comportement de M-LWDF est ici bien moins satisfaisant, puisque les contraintes de QoS ne sont entièrement respectées que pour les deux derniers MODCOD. Une fois de plus, on constate de grandes différences de délai entre les MODCOD, et entre les différentes QoS d'un même MODCOD. Ce résultat montre le manque de flexibilité des solutions empiriques, dont l'adaptation à un scénario ne présuppose pas la performance dans d'autres.

Le cas de UBMT est plus intéressant : ses performances semblent similaires au scénario précédent, bien qu'encore une fois en dehors des contraintes de QoS définies. Cette propriété semble intéressante, même si nous avons mesuré que, dans notre cas, moins de 1% des MODCOD sont sélectionnés par la partie Urgency-Based de l'algorithme, ce qui limite son intérêt, étant donné la complexité de l'étape. De plus, la gestion de critères fins de QoS, comme ceux que nous avons définis, n'est pas facilitée par UBMT, du moins dans les versions présentées dans [5] et [61]. Les travaux sur UBMT n'ont pas été poursuivis, à notre connaissance, mais il semble qu'une étude plus approfondie de cette famille d'algorithmes pourrait permettre une amélioration significative.

Enfin, l'algorithme BPS semble offrir des performances quasiment identiques à celles du scénario précédent. On ne mesure que peu de différences entre les MODCOD, et entre les différentes QoS. Cette propriété, validée ainsi dans un cas très dégradé, conforte les premiers résultats obtenus dans le scénario ciel clair, quant à la flexibilité de BPS. De plus, nous avons observé des résultats similaires, entre BPS et BBS, sinon légèrement meilleurs pour BBS, au détriment d'une complexité accrue.

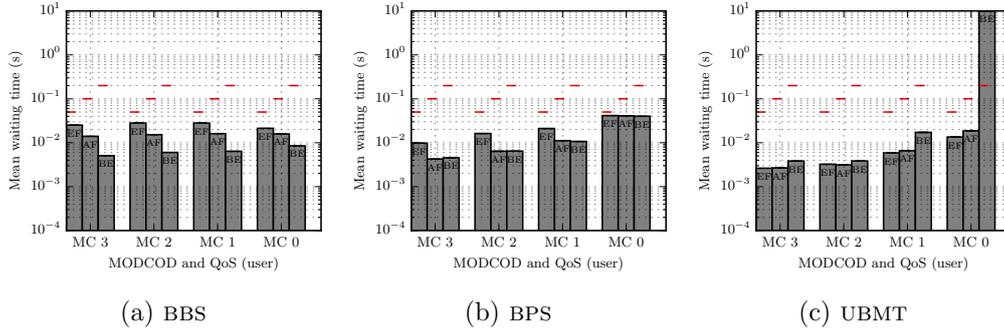


FIGURE 3.6 Résultats de simulation pour le scénario «iso-ModCod»

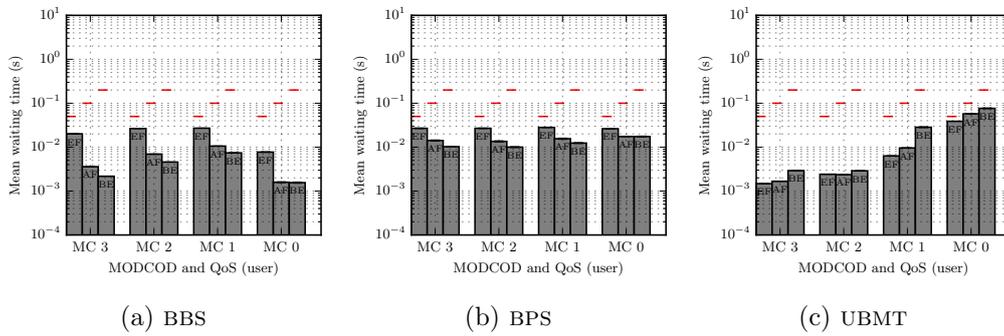


FIGURE 3.7 Résultats de simulation pour le scénario «iso-QoS»

3.5.4 Résultats scénarios d'analyse

Les trois algorithmes retenus pour cette analyse sont BBS, BPS et UBMT.

Les résultats pour le scénario «iso-MODCOD» sont présentés sur la figure 3.6. On peut observer pour ce scénario un comportement des algorithmes comparable à ce que nous avons déjà observé, à savoir une performance très similaire pour BBS et BPS. Nous pouvons de plus noter que UBMT semble ne pas pouvoir ordonnancer correctement le trafic BE du MODCOD le moins fort (MC0). Ce résultat montre l'indépendance de la performance de notre algorithme à la répartition de la charge entre les MODCOD. Les performances d'UBMT peuvent s'expliquer par la conjonction de la faible efficacité spectrale (faible débit) du MODCOD 0 et l'absence de prise en compte d'un délai maximal pour la classe BE. Le trafic de cette file n'est pas servi suffisamment souvent, comme le montre son délai moyen important.

Les résultats du scénario «iso-QoS» sont présentés dans la figure figure 3.7. Ce scénario mesure l'impact d'une charge équitablement répartie entre les trois QoS. Si des différences apparaissent cette fois entre BBS et BPS, celles-ci sont faibles, de l'ordre de

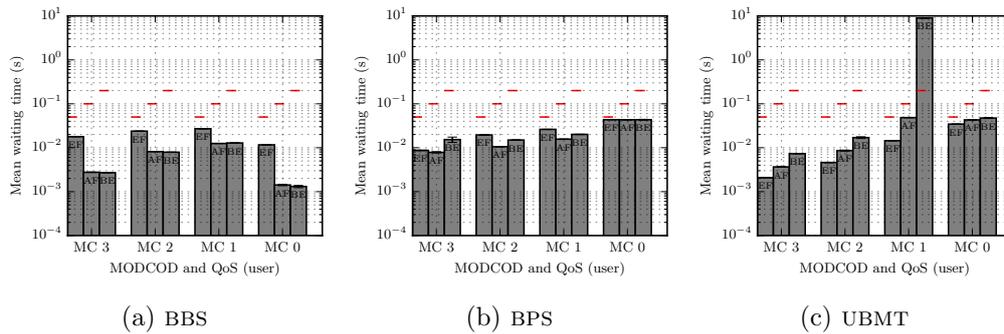


FIGURE 3.8 Résultats de simulation pour le scénario «téléphonie»

quelques millisecondes, par rapport aux critères de QoS définis. Cependant, les différences de performance entre les différentes QoS apparaissent plus distinctement au sein d'un MODCOD, en particulier pour l'algorithme optimal et UBMT. Si ces différences ne sont pas préjudiciables, puisque tous les algorithmes respectent les contraintes de QoS fixées, elles montrent que le comportement limite de l'algorithme peut parfois ne pas être équitable. Un point intéressant est le fonctionnement de BPS, qui semble lui peu affecté par cette iniquité, grâce au lissage du DRR.

Les résultats de notre dernier scénario sont présentés dans la figure 3.8. Nous observons dans ce scénario deux propriétés importantes. La première est que les trois algorithmes présentés ici sont tout à fait capables de fonctionner dans un scénario très éloigné de celui que nous avons fixé, puisque le trafic majoritaire (EF) est ordonnancé conformément aux contraintes de QoS. Si le trafic BE du MODCOD 1 pour UBMT semble ne pas respecter ces dernières, il correspond à un volume faible de données, dans un système où le trafic de ce type n'est généralement pas garanti. Une fois encore, BPS présente une performance homogène et compatible avec la QoS requise, ce qui conforte notre choix de cet algorithme en tant que candidat pour une implantation dans un système réel, du fait de ses nombreux avantages.

3.5.5 Complexité et discrétisation

Au regard de notre objectif d'implantation, déterminer le coût, en terme de calculs, de notre algorithme est un facteur déterminant. Pour ce faire, nous utilisons deux métriques : la complexité dans le pire cas, et le temps de calcul moyen. Si la complexité représente une borne supérieure, clairement définie, elle ne donne pas d'informations sur le comportement moyen de l'algorithme. En revanche, le temps moyen de calcul permet d'observer, pour un matériel et une implantation donnée, la performance de l'algorithme.

Algorithme	Complexité
UBMT [5]	$o(MQ)$
WRR [4]	$o(1)$
M-LWDF	$o(MQ + M + Q \log Q)$
BBS	$o(MQ + M + Q \log Q)$
BPS	$o((MQ)^2)$

TABLE 3.4 Complexité d'algorithmes d'ordonnancement

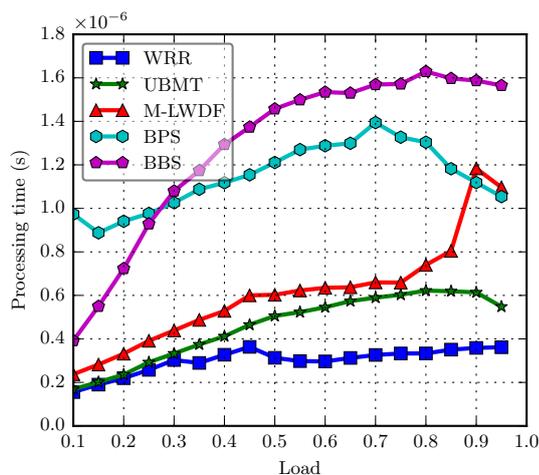


FIGURE 3.9 Temps de calcul par BBFRAME

La table 3.4 donne la complexité des algorithmes que nous avons utilisés. Notons que la complexité d'UBMT est celle de notre adaptation, bien inférieure à celle donnée dans [5], et que celle de BPS ne comporte que l'algorithme d'optimisation, et non le DRR, dont la complexité est $o(1)$. Sans surprise, les algorithmes les moins complexes, comme WRR sont ceux présentant la complexité la plus basse, la complexité la plus haute revenant à BPS. Ce résultat peut sembler paradoxal par rapport à ce qui a été dit plus haut, mais il faut préciser que cette complexité est liée à la périodicité de l'algorithme d'ordonnancement, plus faible pour BBS que pour BPS. Ainsi, la complexité ne nous permet pas ici de déterminer effectivement le compromis entre performance et coût calculatoire, mais est une information indispensable pour prédire le comportement de l'algorithme.

Sur la figure 3.9, le temps moyen de calcul *par* BBFRAME est représenté en fonction de la charge. Bien que ce résultat soit très dépendant de l'implantation effectuée, l'écart relatif entre deux algorithmes n'en est pas moins pertinent, dans la mesure où nous

utilisons le même simulateur, et la même machine de simulation. Les trois algorithmes empiriques, WRR, UBMT et M-LWDF semblent peu sensibles à la charge, et montrent des temps de calcul similaires, excepté à forte charge. Nous observons également que BBS et BPS ont un comportement bien plus marqué par la charge. BPS possède un coût inférieur, à partir d'une charge de 0.25, que BBS, même si la différence est faible, ce qui est relativement décevant étant donné la complexité du calcul effectué pour BPS. Il est possible, bien sûr, de modifier la durée de la période d'ordonnancement pour BPS, et baisser ainsi le coût par BBFRAME, mais on risque d'augmenter l'erreur dans l'estimation de l'état final du système. Il y a donc un compromis à trouver sur cette valeur également, mais elle suppose une étude plus fine du comportement temporel du système.

Un dernier point important est le passage continu-discret, requis par BBS et BPS. Afin de justifier l'utilisation d'algorithmes d'optimisation complexes, ce passage doit non seulement avoir une complexité faible, mais également ne pas dénaturer la solution trouvée de manière trop importante. Nous avons pu déterminer qu'environ 90% du volume ordonnancé par BPS peut être obtenu, à partir de la solution continue, par une simple division entière. Le système doit donc déterminer, à l'aide de méthodes d'arrondis, approximativement 10% du volume ordonnancé, ce qui est relativement faible, mais pas négligeable. Une fois de plus, ce chiffre peut être largement réduit par l'allongement de la période d'ordonnancement, mais avec les mêmes risques pour la précision de la solution que ceux cités précédemment. De plus, l'implantation donnée pour BPS, utilisant un DRR permet de mettre en œuvre cette discrétisation de manière "naturelle" : il suffit d'arrondir les volumes trouvés à l'entier le plus proche, ceux-ci étant utilisés pour le calcul du quantum et du déficit, qui peuvent prendre n'importe quelle valeur entière. La pertinence du passage en réel, afin de déterminer la solution d'ordonnancement, semble donc être assurée, au regard des performances obtenues, et de l'influence effective du passage continu-discret. Notre solution permet également de contourner le problème soulevé par Chaput et al. dans [28], concernant la résolution sous-optimale du Knapsack.

3.5.6 Conclusion sur l'évaluation de performances

L'évaluation de performances menée dans cette partie nous a permis de tirer les conclusions suivantes. Tout d'abord, l'adaptation à notre système des algorithmes existants s'est avérée plus complexe que prévue. En effet, les algorithmes proposés dans la littérature, particulièrement WRR et UBMT reposent sur des caractéristiques très particulières du trafic, et leur adaptation à notre système n'a pu se faire qu'au prix de modifications importantes. Nous avons pu confirmer la difficulté à adapter les paramètres (équité, poids de QoS) à un scénario en particulier. Ce problème peut être ignoré pour un système dont

le scénario change peu, mais cette supposition n'est pas réalisable avec notre système de référence. Ainsi, M-LWDF pourrait être adapté à d'autres types de systèmes, plus stables dans le temps, dans la mesure où il permet d'obtenir de bonnes performances pour un coût calculatoire faible.

Nos algorithmes proposés, BPS et BBS, ont montré leur efficacité et leur capacité d'adaptation. Un intérêt particulier de notre solution est l'homogénéité de ses performances, tant sur des scénarios différents, qu'au sein d'un même scénario. En effet, le fait d'avoir un délai équivalent, indépendamment du MODCOD est un avantage certain pour la couche accès, dont l'influence est ainsi moins visible.

3.6 Conclusion sur l'ordonnancement voie Aller

Notre objectif initial consistait à proposer une solution d'ordonnancement, prenant en compte à la fois le caractère dynamique de la demande et de la capacité, mais également les contraintes de QOS. De plus, l'algorithme devait également être compatible avec les contraintes temps réel du système.

Nous avons ainsi proposé et développé deux types de solutions, l'une reposant sur les règles d'ordonnancement et l'autre sur les fonctions d'utilité. Ces algorithmes permettent d'ordonner des BBFRAME, c'est-à-dire de déterminer à la fois le MODCOD et les paquets encapsulés, en respectant les contraintes de capacité.

Nous avons proposé deux méthodes de conception : l'une pour les règles d'ordonnancement, l'autre pour les fonctions d'utilité. Si la première est simple, et ne laisse que peu de choix quant à l'implantation effective de l'algorithme, ce n'est pas le cas des règles d'ordonnancement, pour lesquelles un algorithme doit être spécifié.

Si les règles d'ordonnancement permettent d'obtenir une solution simple et peu coûteuse en calculs, elles se sont révélées trop sensibles au scénario pour notre application. L'absence d'objectif à long terme fixé rend, de plus, ces solutions difficiles à maîtriser.

L'algorithme BPS semble fournir une réponse efficace au problème fixé en début de chapitre, tant par ses performances, sa prise en compte explicite de la QOS, que son temps de calcul par BBFRAME, moins élevé que BBS. Il permet d'évaluer de façon claire chaque solution d'ordonnancement selon une métrique commune, et nous avons pu observer que son comportement est relativement similaire d'un scénario à l'autre. Cette dernière propriété est importante : si le comportement de la couche accès est similaire dans beaucoup de scénarios, alors le comportement des couches supérieures devient lui-même plus facile à analyser. Les protocoles de transport, notamment, seront plus efficaces dans un environnement où le délai de transmission de bout en bout est stable. Ces avantages

en font une solution répondant aux besoins exprimés pour l'ordonnancement sur la voie Aller. De plus, la faible complexité du DRR est un avantage certain pour une implantation dans un environnement réel.

ALGORITHMES D'ALLOCATION DE RESSOURCES POUR LA VOIE RETOUR

Dans ce chapitre, nous présentons une solution pour l'allocation de ressources sur la voie Retour adaptée à un environnement DVB-RCS2. Comme nous avons pu le voir dans le Chapitre 2, les solutions proposées dans la littérature sont moins nombreuses que pour la voie Aller. Si les mêmes outils (critères explicites, fonctions d'utilité et théorie des jeux) ont pu être utilisés, l'architecture spécifique de la voie Retour influence grandement les solutions d'allocation de ressources, principalement du fait du comportement bouclé de l'allocation.

Du fait des avantages indéniables des outils retenus pour construire nos algorithmes sur la voie Aller, nous proposons de réutiliser critères explicites et fonctions d'utilité pour la voie Retour ; disposer d'une formalisation similaire apportera une meilleure compréhension du problème. Notre objectif, dans ce chapitre, est d'apporter une réponse à la question de l'allocation de ressources, c'est-à-dire de fournir un algorithme donnant, en fonction de l'état du système (requêtes, MODCOD) une allocation des BTU. Cette allocation doit se faire de manière équitable, tout en respectant la QoS, et ayant un temps de calcul compatible avec des contraintes temps réel. De plus, les problèmes de congestion (au sens de l'allocation de ressources) et de FCA, généralement traités séparément, devront être intégrés au même processus, afin de disposer d'un algorithme unifié. Cette unification permettra un comportement cohérent du système, quelque soit son état.

Deux problèmes majeurs se posent pour la voie Retour : l'estimation des requêtes de trafic, et l'allocation de ressources. Si l'allocation de ressources peut être traitée, comme nous allons le montrer, de façon très similaire à la voie Aller, il n'en est pas de même pour l'estimation des requêtes. En effet, cette dernière suppose non seulement une étude

approfondie des profils de trafic, mais aussi de leur comportement dans un système de type DAMA. De nombreux travaux ont été consacrés à ce problème [70] [72] [19] [81], montrant son intérêt. En particulier, les outils utilisés pour l'analyse de ces algorithmes sont bien différents de ceux que nous avons utilisés jusqu'ici (boucles de contrôle, évaluation analytique de performances). Aussi, l'étude de l'estimation des requêtes constitue une thématique de recherche bien distincte, différente de celle que nous avons traitée jusqu'ici, et requerrait une étude à part entière. Dés lors, dans l'objectif d'optimiser l'allocation de ressources en particulier, il est nécessaire de l'isoler en considérant les autres processus comme fixés. Nous considérerons ainsi une solution d'estimation des requêtes, « la meilleure possible », considérant l'état des recherches actuelles.

Ce chapitre sera composé de trois parties. Dans la première, nous présentons en détail l'architecture de notre solution, les paramètres et les entrées/sorties de notre algorithme d'allocation de ressources, et l'algorithme de référence que nous utilisons pour comparer les performances. Dans une deuxième partie, nous proposons une solution s'appuyant sur les fonctions d'utilité. L'évaluation des performances, effectuée à partir de simulations, est présentée dans la dernière partie.

4.1 Architecture du système

Nous présentons dans cette section les paramètres décrivant notre système de référence, ainsi que les mécanismes d'allocation considérés. Notre système de référence pour la voie Retour est tel que décrit au Chapitre 1.

4.1.1 Paramètres et boucle d'allocation

Nous considérons un ensemble de N Terminaux, indexés par la variable i , ayant chacun Q classes de service, de type DIFFSERV, indexées par la variable j . La Supertrame comporte L par K BTU (L selon l'axe fréquentiel, K selon l'axe temporel, cf figure 1.6), où $M = KL$ BTU au total.

Le temps est mesuré de façon discrète, et indexé par le numéro de la Supertrame DVB-RCS2, n . Toutes les durées seront ainsi données en nombre équivalent de Supertrames.

Un schéma de la boucle d'allocation pour un Terminal est donné sur la figure 4.1. A chaque Supertrame, un sous-ensemble de Terminaux possède un SYNC SLOT pour transmettre ses requêtes de trafic, et la mesure du MODCOD optimal, notées respectivement x_{req_i} et m_i . Notons qu'il est également possible, pour un Terminal disposant déjà d'une allocation, d'envoyer cette requête conjointement avec ses données [16]. Chaque classe de QoS choisit un type de requête, RBDC, VBDC ou bien utilise le CRA. Cette requête est

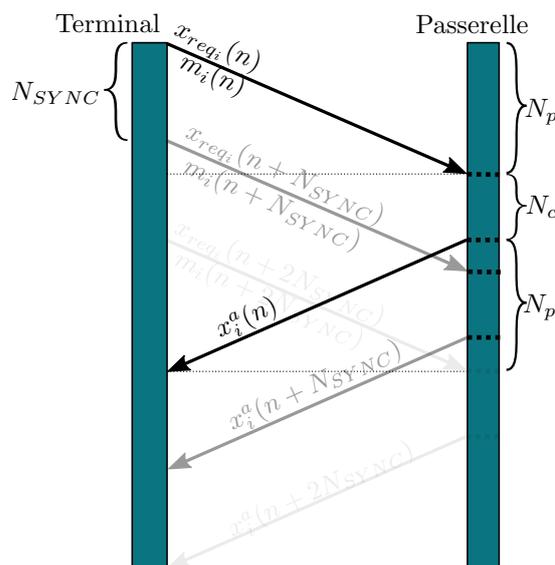


FIGURE 4.1 Chronogramme de la boucle d'allocation

notée $x_{req_i,j}$ pour la QOS j . On notera $X_{req} (\in \mathbb{N}^N \times \mathbb{N}^Q)$ la matrice des requêtes et $m (\in \mathbb{N}^N)$ le vecteur des MODCOD.

Nous choisissons d'exprimer X_{req} en nombre équivalent de BTU, ce qui, d'un point de vue système n'est pas tout à fait exact (mais a un impact négligeable), puisque la correspondance est normalement faite sur la Passerelle plutôt que par le Terminal. Ainsi, le débit effectif pour une BTU sera dépendant du MODCOD, puisque sa taille est donnée en nombre de symboles. C'est ici un avantage de DVB-RCS2, simplifiant l'expression du problème d'allocation de ressources en unités de même taille.

Une fois ces données collectées, au bout d'un temps de propagation de N_p Supertrames, la Passerelle dispose de N_c Supertrames pour calculer l'allocation correspondante, x_i^a . Par convention, on désignera par $x_{req_i}(n)$ la requête (comprenant toutes les QOS) envoyée à l'instant n par le Terminal i , qui sera donc reçue par la Passerelle à l'instant $n + N_p$. Ainsi, l'allocation correspondant à la requête $x_{req_i}(n)$ sera reçue à $n + 2N_p + N_c$. Cette allocation, notée $x_i^a(n + N_p + N_c)$ est unique par Terminal (même si plusieurs requêtes sont fournies). Le vecteur des allocations sera noté X^a

Du point de vue de la Passerelle, pour chaque nouvelle Supertrame n , les variables reçues des Terminaux sont $X_{req}(n - N_p)$ et $m(n - N_p)$. L'algorithme d'allocation de ressources devra donc prendre en compte, à chaque nouveau calcul, ces paramètres. C'est ici que le caractère très particulier de l'allocation de ressources DVB-RCS2 se fait ressentir :

contrairement aux systèmes LTE, où l'échelle temporelle du processus est de quelques millisecondes, celle de DVB-RCS2 est d'au moins deux temps de propagation, soit une demi-seconde. Si cette période n'est pas directement dépendante du temps de propagation, elle en est tributaire, car elle limite la période minimale de mise à jour : l'allocation, faite ici toutes les Superframes, pourrait être plus lente. Ainsi, la prise en compte du caractère dynamique dans DVB-RCS2 est complexe, car il ne peut se faire qu'au travers des requêtes, ou bien de métriques évaluées sur la Passerelle, qui ne donnent qu'un aperçu retardé du système.

Ce comportement bouclé, avec retard, aurait pu nous mener vers des solutions bien différentes que les fonctions d'utilité. En effet, un certain nombre de travaux considèrent, pour évaluer ces systèmes, une approche utilisant une boucle de contrôle [81]. Cette approche, ici ne nous semble pas adaptée au seul problème d'allocation de ressources, puisqu'elle prend généralement en compte également une estimation du trafic. De plus, les résultats des prédictors de trafic sous un retard important sont particulièrement disparates [70] [82].

4.1.2 Propriétés du trafic voie Retour

Dans un système de type *broadband*, comme celui étudié ici, le trafic engendré par les Terminaux sur la voie Retour possède des caractéristiques bien différentes de celui de la voie Aller.

En effet, si le visionnage de vidéo, le transfert de fichiers ou encore la VOD demandent un débit important sur la voie Aller, leur pendant sur la voie Retour est constitué majoritairement d'accusés de réception Transmission Control Protocol (TCP), et engendre donc un trafic très faible. Les seules applications véritablement exigeantes en termes de débit ne sont pas courantes : vidéoconférence, *upload* de vidéo ou de fichiers. La tendance actuelle de stockage à l'intérieur du réseau (« dans le nuage ») affirme d'autant plus cette dissymétrie. L'International Telecommunication Union (ITU), par l'intermédiaire du 3GPP2, définit dans [83] un modèle de trafic, pour l'évaluation des réseaux mobiles, confirmant cette propriété.

Ainsi, si nous conservons la classification DIFFSERV adoptée jusqu'ici pour le trafic, la répartition de la charge entre chaque QOS sera bien différente. Nous supposons également que le Terminal utilise majoritairement les requêtes RBDC et VBDC, le CRA étant réservé au trafic de signalisation. Nous supposons également qu'un même terminal peut différencier ses requêtes grâce au `rc_index` défini dans [16], qui permet de différencier la priorité de deux requêtes provenant d'un même Terminal, autrement que par la nature de la requête (RBDC,VBDC).

Nous considérons, pour l'algorithme d'allocation, une requête exprimée en équivalent en nombre de BTU, suivant trois classes de QOS, EF (n°0) AF (n°1) et BE (n°2). Le type de requête est RBDC pour les deux premières QOS, EF et AF, et VBDC pour BE. Le Terminal recevant une allocation unique, l'ordonnancement est laissé à son initiative.

4.1.3 Algorithme de référence

Nous présentons ici un algorithme de référence, dérivé de [71], apportant une réponse simple, sous forme d'un critère explicite, aux critères que nous avons fixés : gestion de la QOS, de la congestion et du FCA dans un même algorithme, allocation équitable, et faible complexité.

L'algorithme défini dans [71] ne permet pas explicitement de prendre en compte la QOS. Il consiste simplement à multiplier chaque requête par un facteur d'échelle, dépendant du niveau de congestion. Aussi, nous définissons, pour la QOS j_1 , le coefficient σ_{j_1} :

$$\sigma_{j_1} = \min \left(\frac{C - \sum_{j=0}^{j_1-1} \sigma_j \sum_i x_{req_{i,j}}}{\sum_i x_{req_{i,j_1}}}, 1 \right) \quad (4.1)$$

Autrement dit, le calcul du coefficient est similaire à [71], excepté qu'il se fait sur la capacité restante, après avoir alloué la capacité nécessaire aux QOS les plus prioritaires ($j < j_1$). Pour un trafic BE, qui est généralement le moins prioritaire, on pourra enlever l'opérateur min, afin de lui allouer la capacité résiduelle. L'allocation totale du Terminal i sera alors :

$$x_i^a = \sum_j \sigma_j x_{req_{i,j}} \quad (4.2)$$

Cet algorithme simple, appelé *linéaire*, nous permettra de disposer d'une référence dans l'évaluation des performances des algorithmes que nous proposons. Il peut être vu comme un critère explicite très simple, une priorité fixe par QOS et une allocation des ressources résiduelles, le cas échéant, aux requêtes correspondant à des flux élastiques. De plus, cet algorithme ne prend en compte qu'un nombre équivalent de BTU (et non un débit), ainsi l'allocation résultante, pour les Terminaux, sera fonction du MODCOD. En considérant que les Terminaux ont des besoins en trafic similaires, un Terminal ayant un mauvais MODCOD demandera, pour le même débit, plus de ressources qu'un Terminal ayant un bon MODCOD. Ainsi, cet algorithme risque d'être inéquitable et privilégier les Terminaux ayant un mauvais MODCOD.

4.2 Algorithme d'allocation de ressources optimisé

Dans cette partie, nous présentons un algorithme d'allocation de ressources faisant usage de fonctions d'utilité, d'après une méthode proche de celle proposée dans le chapitre précédent pour la voie Aller. Cette formulation doit permettre de formuler le problème d'allocation de ressources de manière synthétique, de façon à éviter la coexistence de plusieurs algorithmes (gestion de la congestion et de la ressource non utilisée) dans le système. Notre objectif, une fois de plus, consiste à fournir une solution dans un contexte exigeant, où la décision d'ordonnancement est soumise à des contraintes temporelles fortes, et où la puissance de calcul, bien qu'importante, est limitée.

Nous présentons tout d'abord la formulation mathématique de notre problème, puis l'implantation qui en a été faite, ainsi que les caractéristiques de l'algorithme finalement obtenu.

4.2.1 Formalisation

L'architecture du système DVB-RCS2 laisse peu de latitude quant au choix des métriques qui vont guider la décision. En effet, bien qu'il soit possible de faire usage de métriques moyennes (mesurées sur la Passerelle), c'est-à-dire dont la valeur dépend du passé, leur pertinence n'est pas assurée. En effet, l'estimation des requêtes de trafic est un processus utilisant les valeurs passées pour prédire la demande future [70]. L'allocation de ressources doit certainement prendre en compte les requêtes, mais doit-elle prendre également en compte les allocations passées ?

La question est bien plus complexe que sur la voie Aller, car les processus régissant la performance du système sont distribués sur les Terminaux et la Passerelle, qui ne pourront prendre une décision cohérente en même temps, du fait du temps de propagation et du délai de mise à jour des requêtes. En effet, comment justifier que, lorsqu'un Terminal met sa requête à jour lors d'un SYNC SLOT, l'allocation correspondante dépende également de celle du SYNC SLOT précédent, qui peut correspondre à un trafic bien différent ? Nous présentons donc une version «sans mémoire» de notre algorithme.

Aussi, durant toute notre discussion, omettrons-nous l'indice temporel n , considérant que nous travaillons avec les requêtes et MODCOD disponibles à l'instant d'ordonnancement n , envoyés N_p Supertrames auparavant.

Fonctions d'utilité

L'utilité d'un Terminal est l'utilité apportée par chacune de ses allocations. Cette utilité prend, en paramètre, les requêtes X_{req} , les MODCOD, ainsi que la QOS. L'objectif ici est de mesurer la satisfaction de l'utilisateur i en fonction de l'allocation reçue, sa correspondance aux contraintes de QOS, tout en prenant en compte les requêtes des autres Terminaux. La prise en compte du MODCOD est nécessaire pour rendre compte de l'efficacité spectrale dans l'allocation de ressources.

Étant donné que l'allocation x_i^a est globale à toutes les QOS, nous définissons $x_{i,j}$ l'allocation de la QOS j pour le Terminal i , ainsi que X la matrice des $x_{i,j}$. Soit $u_{i,j}$ la fonction d'utilité associée. Cette fonction dépend de trois paramètres : l'allocation $x_{i,j}$, la requête et le MODCOD. Cependant, travaillant à n fixé, $x_{req_{i,j}}$ ainsi que m_i sont fixés. L'utilité est donc fonction de la seule allocation $x_{i,j}$, à n fixé. Nous pouvons donc utiliser, sans ambiguïté, la notation suivante :

$$u_{i,j}(x_{i,j})$$

pour désigner la fonction d'utilité. Cette fonction sera paramétrée, pour chaque Suprame n , par :

- la requête $x_{req_{i,j}}$;
- le MODCOD m_i ;
- la QOS j .

Nous avons proposé la fonction suivante [26] :

$$u(x_{i,j}) = \alpha_{i,j}x_{i,j} + (1 - \exp \beta_{i,j}(x_{i,j} - x_{req_{i,j}}))$$

Où :

- $\alpha_{i,j}$ est un paramètre dépendant du MODCOD de l'utilisateur i et de la QOS ;
- $\beta_{i,j}$ est un paramètre de QOS.

On distingue deux cas dans le comportement de cette fonction :

- Si $x_{i,j} < x_{req_{i,j}}$ alors la valorisation est exponentielle, de paramètre $\beta_{i,j}$;
- Si $x_{i,j} > x_{req_{i,j}}$ alors la valorisation est linéaire, de paramètre $\alpha_{i,j}$.

On trouvera un compromis dans les valeurs de α et β : pour les flux ayant une QOS élevée, β sera important, et l'utilité diminuera donc fortement avec x . Cependant, ces flux à forte QOS sont généralement non élastiques, il n'y aura donc que peu d'avantages à les servir au-delà de x_{req} , d'où une valeur faible de α (voire zéro pour les flux de type CBR). Au contraire, les flux élastiques auront une QOS assez basse (d'où un β faible), mais

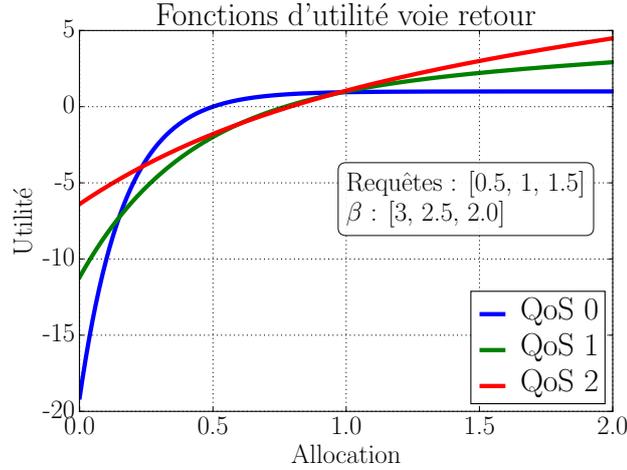


FIGURE 4.2 Fonctions d'utilité pour la voie Retour

pourront profiter d'une bonne qualité du support (d'un fort MODCOD) pour augmenter leur débit, d'où une valeur de α plus importante.

Pour α , nous proposons la valeur suivante :

$$\alpha_{i,j} = a_{i,j} \left(1 + \frac{c(m)}{\sum_m c(m)} \right)$$

Où $c(m)$ est l'efficacité spectrale associée au MODCOD m , utilisé par le Terminal i , la somme étant faite sur tous les MODCOD utilisés dans le système. $a_{i,j}$ vaut zéro pour les flux inélastiques, 1 pour les flux vidéo, et 2 pour les flux temps réel. Ce paramètre rend compte du compromis entre efficacité spectrale et QoS, dans le cas où la capacité est suffisante pour être en FCA. Le comportement de cette fonction est illustré sur la figure 4.2.

4.2.2 Problème d'optimisation

Nous formulons l'allocation de ressources sous forme d'un problème d'optimisation, qui doit être résolu à chaque Supertrame. Étant donné notre choix de système, et contrairement à la voie Aller, il est nécessaire de résoudre exactement ce problème, et non de trouver la direction de la solution optimale, celle-ci doit être atteignable, faute de quoi l'allocation n'est pas réalisable. Dans la suite, nous supposons qu'il existe au moins une allocation réalisable¹.

1. Cette condition peut être réalisée par des contraintes sur les valeurs extrêmes que l'allocation peut prendre

Formulation

L'objectif du système est, pour chaque Supertrame, de maximiser l'utilité de tous les Terminaux. Nous pouvons exprimer cette utilité de la façon suivante :

$$U(X) = \sum_{i=0}^{N-1} \sum_{j=0}^{Q-1} u_{i,j}(x_{i,j})$$

Afin de simplifier les expressions, nous noterons $\sum_{i,j}$ la somme précédente. Le problème d'optimisation à résoudre, pour chaque Supertrame, sera donc le suivant :

$$\begin{aligned} & \underset{X}{\text{maximiser}} && U(X) = \sum_{i,j} u_{i,j}(x_{i,j}) \\ & \text{avec} && \sum_{i,j} x_{i,j} = M \\ & \text{et} && \forall i \sum_j x_{i,j} \leq K \\ & \text{et} && \forall i \forall j x_{\min_{i,j}} \leq x_{i,j} \leq x_{\max_{i,j}} \end{aligned} \tag{4.3}$$

La première contrainte représente la capacité du système : on ne peut pas allouer plus que le nombre de BTU disponible. La deuxième exprime la faisabilité de l'allocation de ressources : les Terminaux Satellite ne peuvent émettre sur deux fréquences en même temps, cette contrainte pose une condition sur le nombre maximum de BTU qu'un Terminal peut se voir allouer, qui ne doit pas dépasser le nombre de BTU par porteuse de la Supertrame. Cette condition nécessaire et suffisante assure la faisabilité de l'allocation de ressources. La dernière contrainte représente des bornes pour l'allocation, dépendantes de la QOS et dont la valeur est fixe pour un Terminal donné. Nous noterons \mathcal{X} le polyèdre formé par cette dernière contrainte.

Résolution

La résolution de ce problème ne peut se faire qu'en considérant les composantes de X comme réelles, et non entières. En effet, dans ce dernier cas, le problème 4.3 est NP-difficile [58] [76], et ne donc peut être résolu en temps réel. Cette hypothèse est fautive d'un point de vue système, puisque l'allocation ne peut se faire qu'en nombre entiers de BTU. Il nous faudra donc spécifier le passage de cette allocation réelle à l'allocation entière, et son influence. Le problème 4.3 appartient à la catégorie des problèmes d'optimisation non linéaires, convexes, et à contraintes linéaires. Étant donné que u est strictement concave, l'ensemble des solutions est convexe, les contraintes sont toutes affines, on peut affirmer

qu'il existe une unique solution au problème 4.3 (conditions de Karush–Kuhn–Tucker [76]).

Soit $L(X, \lambda, \mu)$ le Lagrangien associé au problème 4.3 :

$$L(X, \lambda, \mu) = \sum_{i,j} u_{i,j}(x_{i,j}) + \lambda(M - \sum x_{i,j}) + \sum_i \mu_i(K - \sum_j x_{i,j}) \quad (4.4)$$

Les conditions d'optimalité de premier ordre s'écrivent :

$$\frac{\partial}{\partial X} L(X, \lambda, \mu) = 0 \Leftrightarrow \sum_{i,j} \frac{\partial}{\partial x_{i,j}} u_{i,j}(x_{i,j}) - \lambda - \sum_i \mu_i = 0 \quad (4.5)$$

avec $X \in \mathcal{X}$. Nous pouvons observer que 4.5 est séparable, comme indiqué dans [76]. Nous pouvons alors écrire l'allocation optimale X^* , dont les composantes sont :

$$x_{i,j}^* = \left[x_{req_{i,j}} - \frac{1}{\beta_{i,j}} \ln \left(\frac{\alpha_{i,j} + \lambda + \mu_i}{\beta_{i,j}} \right) \right]_{x_{min_{i,j}}}^{x_{max_{i,j}}} \quad (4.6)$$

où $x = [y]_a^b$ est la projection du résultat sur l'ensemble \mathcal{X} .

Nous pouvons remarquer d'après 4.6 que $x_{i,j}^*$ est une fonction de λ et μ_i . Il suffit alors de déterminer leur valeur pour obtenir X^* . Une résolution possible, proposée par [29], consiste à utiliser le problème dual, équivalent à 4.3.

$$\min_{\lambda, \mu} D(\lambda, \mu)$$

Où

$$D(\lambda, \mu) = \max_X L(X, \lambda, \mu)$$

Ce problème peut alors être résolu par une méthode itérative [76] de projection du gradient de λ et μ :

$$\lambda(k+1) = \left[\lambda(k) - \gamma \left(M - \sum_{i,j} x_{i,j}^*(\lambda(k), \mu_i(k)) \right) \right]^+$$

$$\mu_i(k+1) = \left[\mu_i(k) - \gamma \left(K - \sum_j x_{i,j}^*(\lambda(k), \mu_i(k)) \right) \right]^+$$

Où γ est un facteur d'amortissement inférieur à 1. Il est possible d'augmenter l'efficacité de cette méthode en prenant, pour γ , les valeurs données par l'algorithme de Newton, qui utilise le gradient afin d'accélérer la convergence.

4.2.3 Implantation

L'algorithme d'allocation de ressources découlant du problème 4.3 nous donne une allocation optimale, mais non réalisable car effectuée en nombres réels. C'est également le cas de l'algorithme linéaire. Nous donnons donc ici une méthode de discrétisation applicable aux deux algorithmes.

L'étape de discrétisation est ici très importante, puisqu'elle peut potentiellement concerner la majorité des Terminaux : considérant les modèles de trafic, la requête dépasse rarement quelques BTU.

L'allocation optimale du Terminal i , x_i^{a*} s'exprime ainsi :

$$x_i^{a*} = \sum_j x_{i,j}^*$$

Ainsi, étant donné que la partie fractionnaire de x_i^{a*} est bornée, son importance réduit lorsqu'on somme les allocation optimales. Cette étape n'a pas d'influence sur le résultat final, puisque l'allocation envoyée au Terminal ne différencie pas les QoS. Ainsi, l'allocation entière qu'on peut déduire directement de X^* est :

$$\lfloor x_i^{a*} \rfloor = \lfloor \sum_j x_{i,j}^* \rfloor$$

La capacité résiduelle, une fois cette allocation faite est :

$$M - \sum_i \lfloor x_i^{a*} \rfloor$$

Cette capacité résiduelle doit être allouée aux terminaux, si possible en fonction de la valeur de $\{x_i^{a*}\}$ (partie fractionnaire de x_i^{a*}), afin de ne pas perdre le bénéfice de l'optimisation de l'allocation réalisée. Aussi, il paraît raisonnable d'allouer la capacité résiduelle dans l'ordre des $\{x_i^{a*}\}$ décroissant. Ainsi, l'allocation finale se rapprochera-t-elle de l'allocation optimale X^* . Il reste cependant à vérifier que ce processus reste peu important en volume par rapport au processus entier, sans quoi la pertinence des calculs complexes permettant de trouver X^* s'en trouverait remise en cause. Nous définissons donc l'efficacité entière de l'allocation comme suit :

$$\eta_e = \frac{\sum_i \lfloor x_i^{a*} \rfloor}{M} \quad (4.7)$$

La discrétisation proposée nécessite cependant de trier les $\{x_i^{a*}\}$, ce qui peut être relativement long pour un grand nombre de Terminaux. De plus, l'allocation reçue change peu

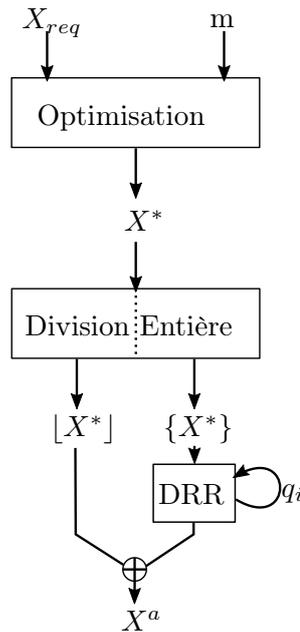


FIGURE 4.3 Algorithme d'allocation de ressources

entre deux SYNC SLOTS, d'où l'idée de cumuler le déficit de BTU sur plusieurs Superframes, et d'attendre que celui-ci soit suffisamment grand pour allouer une BTU supplémentaire à l'utilisateur i . Bien que ce mécanisme biaise quelque peu le processus d'allocation, qui n'est plus indépendant des allocations passées, il permet d'utiliser des algorithmes tels que DRR, bien moins complexes que les algorithmes de tri. Nous avons alors $q_i = \{x_i^{a*}\}$, le quantum du Terminal i , utilisable dans un algorithme de DRR, qui s'occupera de l'allocation de la partie de la Superframe qui n'a pu être déterminée à l'aide de la première étape. Un schéma de cet algorithme est présenté à la figure 4.3.

Nous disposons ainsi d'un algorithme de complexité $o(1)$ pour le DRR, répondant aux critères fixés au début de cette étude pour la voie retour. Il reste cependant à valider son comportement, particulièrement en ce qui concerne l'efficacité entière, et le respect des critères de QoS.

4.3 Évaluation de performances

Dans cette partie, nous détaillons la méthode d'évaluation de l'algorithme *optimal* présenté, ainsi que de l'algorithme de référence. Nous définissons tout d'abord un scénario de référence, correspondant au système décrit dans le chapitre 1. Nous présentons ensuite les résultats des simulations menées.

Cette étape a pour but de montrer, non seulement la faisabilité de notre algorithme, mais également sa performance par rapport à l'algorithme de référence, dit *linéaire*, présenté plus haut.

4.3.1 Scénario de référence

La définition de notre scénario de référence comporte la répartition des MODCODs, selon un scénario de type *ciel clair*, présenté sur la table 4.1, et d'un modèle de trafic par Terminal, dont un résumé est donné à la table 4.2. Ces modèles reposent sur ceux donnés par le 3GPP2, décrits dans [83], où trois types de flux sont considérés : un flux VOIP (standard G.729), 20 sessions HTTP et 20 sessions IP Television (IPTV) correspondant à quelques utilisateurs situés derrière le Terminal, engendrant des flux.

Si le flux VOIP correspond exactement au standard, les flux HTTP et IPTV correspondent, pour la voie Retour, à un flux d'accusés de réception TCP, directement déduit des pages envoyées sur la voie Aller : [83] définit un modèle statistique de génération de pages HTML, leur fréquence d'accès et leur taille, il est alors aisé de déterminer les accusés de réception engendrés par la réception de cette page, en prenant en compte le temps de traitement du navigateur. Nous avons ainsi pu concevoir des sources de trafic adaptées, que nous avons implanté dans un simulateur dédié. De plus, afin d'éviter une soudaine augmentation de la charge, l'ensemble des premières demandes de trafic a été étalée sur les 20 premières secondes de la simulation. Nous observerons donc une montée en charge progressive.

#	ModCod	charge
MC0	QPSK 2/3	0.012
MC1	QPSK 4/5	0.362
MC2	8-PSK 2/3	0.353
MC3	8-PSK 4/5	0.268
MC4	8-PSK 5/6	0.005

TABLE 4.1 distribution voie Retour – Ciel clair

Nous pouvons remarquer que la répartition des MODCOD sur la voie retour est différente de celle donnée sur la voie Aller (table 3.2) : les caractéristiques physiques des récepteurs des Terminaux expliquent en partie cette différence. De plus, si seule la répartition relative comptait sur la voie Aller, l'utilisation de modèles précis de trafic oblige ici à modéliser de façon plus réaliste des conditions de transmission.

Afin de pouvoir contrôler la charge de notre système, nous donnons une estimation

Application	VoIP	VoD	Internet
Classe	EF	AF	BE
débit moyen (kbps)	8	0.6585	1.1735

TABLE 4.2 Répartition des classes de QoS sur la voie Retour

de la charge moyenne ρ_a en fonction des paramètres du scénario :

$$\rho_a = \frac{N \sum_k \frac{d_t}{d_m(k)} p(k)}{M} \quad (4.8)$$

Où :

- d_t est le débit moyen total d'un terminal, défini dans la table 4.2 ;
- $d_m(k)$ est le débit du MODCOD k ;
- $p(k)$ est la charge relative du MODCOD k , définie dans la table 4.1.
- N est le nombre de Terminaux
- M est le nombre total de BTU

L'idée de 4.8 est d'exprimer le trafic moyen engendré par l'ensemble des Terminaux, en nombre équivalent de BTU, et de le diviser par la capacité de la Supertrame DVB-RCS2. Nous pouvons ainsi, connaissant d_t , ajuster le nombre de BTU afin d'obtenir une charge moyenne donnée.

Nous observerons principalement le temps d'attente des paquets, c'est-à-dire leur temps passé dans le système, en fonction de la QoS, ainsi que l'efficacité η_e ((4.7)). Nous définissons de plus deux métriques, la charge système ρ_s (valeur moyenne $\bar{\rho}_s$) et la charge réelle ρ_r (valeur moyenne $\bar{\rho}_r$). La première correspond à la proportion de BTU demandées par rapport au nombre de BTU disponibles (c'est-à-dire le dimensionnement de la trame), la deuxième correspond à la proportion de BTU utilisées dans une trame donnée. La durée de notre simulation est fixée à 150 secondes, et la charge moyenne est 0.8 pour 400 Terminaux. La durée de la Supertrame DVB-RCS2 est fixée à 25 ms.

4.3.2 Résultats de simulation

Les résultats de simulation concernant le temps d'attente sont présentés en figure 4.4, sous forme de fonction de répartition. Cette représentation nous permet une bonne analyse du comportement statistique de l'allocation de ressources, et son impact sur le trafic.

Analysant les résultats présentés sur la figure 4.4a, nous pouvons remarquer que la quasi-totalité du trafic EF possède un très faible délai : 99% se situe en-dessous de 50ms.

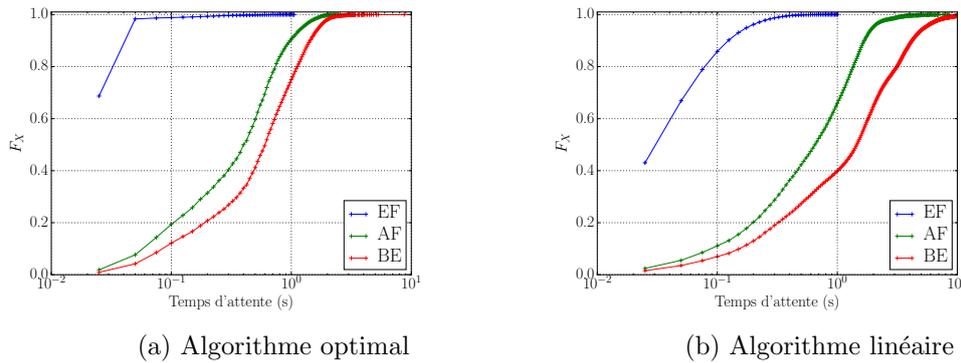


FIGURE 4.4 Fonction de répartition empirique du temps d'attente

Ces performances sont ainsi tout à fait compatibles avec des contraintes strictes de QoS, comme définies au chapitre 3. La performance du trafic AF et EF est bien plus mitigée, puisqu'on remarque qu'une proportion significative du trafic subit un délai de plusieurs secondes. Ce trafic, très irrégulier, aura tendance à créer des *bursts* qui ne seront pris en compte dans l'allocation qu'une fois la requête correspondante émise. Ainsi, un retard d'au-moins deux fois le temps de propagation, en plus du temps de calcul, peut être observé, sans compter le temps de réaction de l'estimateur de requêtes.

Par comparaison avec les résultats de l'algorithme linéaire, présentés sur la figure 4.4b, nous pouvons constater que l'utilisation de notre algorithme permet non seulement de meilleures performances dans l'ensemble des classes de QoS, mais également un meilleur comportement statistique pour AF et BE. En effet, dans le cas de l'algorithme linéaire, ces deux classes subissent une grande variabilité dans le temps d'attente : 60% du trafic BE voit son délai réparti entre 1 et 6 secondes, ce qui peut être particulièrement préjudiciable pour les protocoles de transport comme TCP. Nous notons d'ailleurs une amélioration significative du traitement du trafic BE dans le cas optimal.

Nous pouvons remarquer que, quelque soit l'algorithme, les performances des trafics AF et BE sont mauvaises, ce qui semble indiquer que le problème n'est pas seulement lié à l'algorithme, mais également à la façon dont les requêtes sont effectuées. Cependant, des travaux supplémentaires seraient nécessaires pour avoir une meilleure compréhension de ce phénomène, qui pourraient constituer une suite logique de la présente thèse.

Nous présentons ensuite les résultats de simulations par MODCOD et par QoS, pour le délai, moyennés sur l'ensemble des Terminaux, de manière similaire aux résultats présentés sur la voie aller.

Les résultats de simulation pour le délai sont présentés à la figure 4.5. De manière générale, nous pouvons observer la grande iniquité entre les performances des différentes

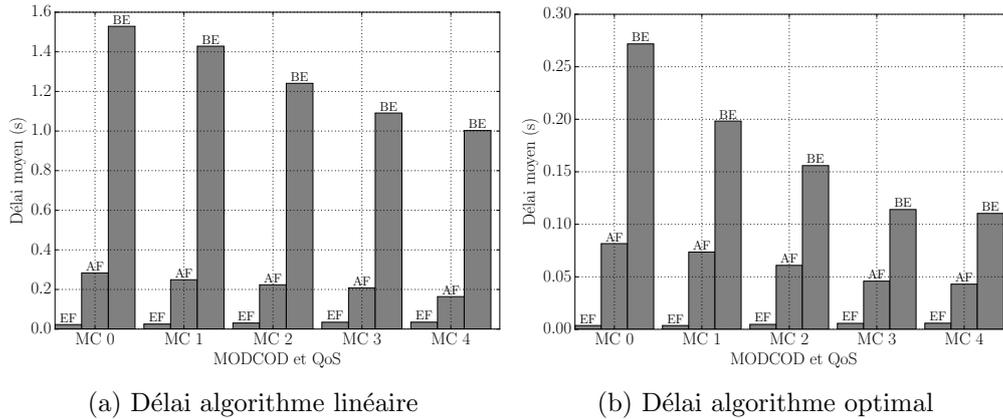


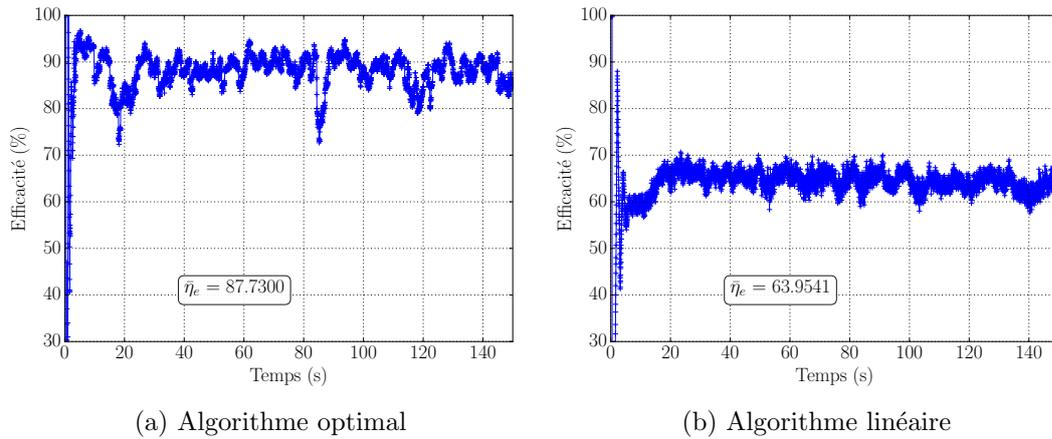
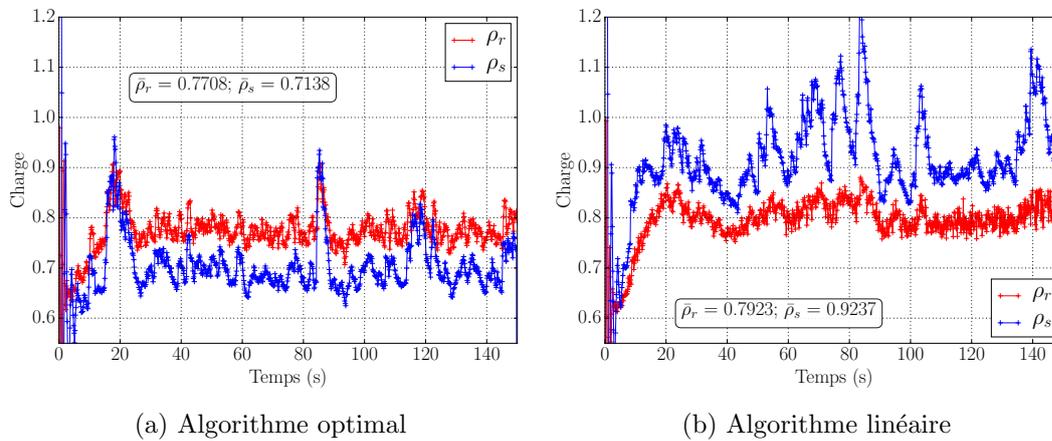
FIGURE 4.5 Résultats de simulation voie retour - délai moyen

QoS au sein d'un même MODCOD, que nous avons déjà remarqué pour le temps d'attente. Si ce résultat n'est pas préjudiciable quand au comportement du système, il montre que la paramétrage des fonctions d'utilité reste complexe.

Nous pouvons remarquer une performance comparable pour chaque MODCOD, manifestement dépendante quelque peu de l'efficacité spectrale : le MODCOD le plus fort (MC4) obtenant la meilleure performance. Si la différence maximale est de 0.5 secondes pour l'algorithme linéaire, elle est de 0.15 secondes environ pour l'algorithme optimal, qui est donc plus équitable au regard de l'efficacité spectrale que notre algorithme de référence. Ce comportement est possible grâce aux fonctions d'utilité que nous avons choisi, qui permettent un compromis entre efficacité spectrale, QoS et requête de trafic.

Enfin, la différence entre l'algorithme optimal et linéaire est ici importante : là où le trafic BE subit un délai moyen de plus de 1.5 secondes avec l'algorithme linéaire, il n'est que de 0.3 secondes, ce qui semble compatible avec une utilisation interactive de type navigation internet.

L'efficacité de l'allocation, présentée à la figure 4.6 montre une différence très marquée entre l'algorithme optimal et l'algorithme linéaire. Alors que l'algorithme linéaire ne dépasse pas 65%, nous observons une efficacité de près de 88% pour l'algorithme optimal. Cela signifie que, comparativement, les allocations sont plus importantes mais accordées à moins de Terminaux dans le cas optimal. L'allocation est mieux répartie dans le cas linéaire, menant ainsi à des allocations plus faibles, et donc une partie plus importante de ces allocations traitées par le DRR. Bien que l'équité soit certainement plus faible dans le cas optimal, nous avons pu voir que celui-ci permet d'écouler le trafic de manière bien plus efficace, remettant en cause la notion d'équité – du moins sur une seule trame – entre les Terminaux. Ce résultat est cependant à nuancer : il dépend de bien d'autres

FIGURE 4.6 Efficacité de l'allocation η_e FIGURE 4.7 Charge système ρ_s et charge réelle ρ_r

paramètres, tels que le nombre de BTU par rapport au nombre de Terminaux, le type de trafic considéré, la façon dont les requêtes de trafic sont construites, etc. Aussi, même si le résultat obtenu est clairement en faveur de notre algorithme optimal, notre seule certitude est que le DRR est marginal dans l'allocation de ressources, et que les calculs complexes de l'optimisation sont efficacement utilisés.

La charge système ρ_s et charge réelle ρ_r sont représentées sur la figure 4.7. Concernant l'estimation de la charge ρ_a , nous pouvons vérifier expérimentalement la pertinence de notre formule : la charge théorique était de 0.8, et nous retrouvons une charge réelle très proche. Cette approximation est, bien sûr dépendante du nombre de Terminaux, ici assez importante pour que l'approximation menée sur les débits moyens soit précise.

Nous notons une différence importante entre l'algorithme linéaire et l'optimal : le

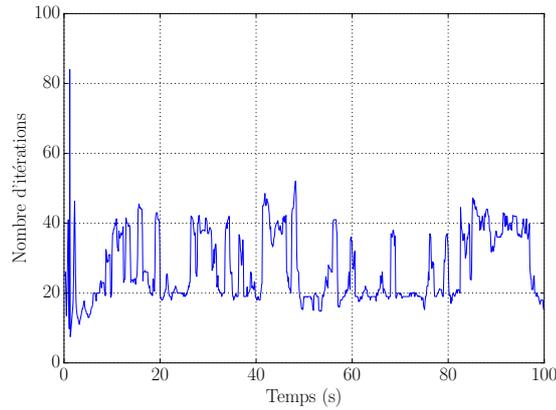


FIGURE 4.8 Nombre d'itérations en fonction du temps

premier montre une plus grande variabilité dans la charge réelle dans les 50 premières secondes. Étant donné la stabilité du cas optimal, nous pouvons en déduire que l'algorithme linéaire aura tendance à créer des instabilités à une échelle de temps assez longue (une dizaine de secondes), un comportement qui pourrait affecter la performance du système.

Concernant la charge système (figure 4.7), nous observons une différence significative : là où l'algorithme linéaire voit sa charge fluctuer largement dans le temps autour d'une valeur plus élevée (0.92), l'algorithme optimal montre une bonne stabilité, mais en sous-estimant quelque peu la charge. Cela signifie que, dans le cas linéaire, les Terminaux demandent plus de BTU que ce dont ils ont effectivement besoin, moins dans l'autre cas. Mis à part quelques *bursts*, dus au trafic, la charge système de l'algorithme montre qu'il est capable de procurer une allocation «lissée» aux Terminaux, correspondant plus à leurs besoins (ce qui confirme les résultats de délai) que l'algorithme linéaire.

Enfin, nous avons mesuré le nombre d'itérations nécessaire à notre algorithme pour converger vers la solution optimale. Ces résultats sont présentés en figure 4.8. Nous pouvons observer que, dans la majorité des cas, l'algorithme converge en moins de 50 itérations, ce qui semble, étant donné la complexité d'une itération, être compatible avec une implantation temps réel.

4.3.3 Conclusion sur l'évaluation de performance

Les travaux d'évaluation de performance menés nous ont permis de valider le fonctionnement de notre algorithme *optimal*. Dans le cas d'un scénario classique, nous avons montré comment il permet une meilleure efficacité, en répartissant de manière plus pertinente les ressources en fonction des besoins des Terminaux, ce qui n'était pas le

cas de l'algorithme linéaire. Nous avons également observé une nette amélioration du traitement moyen des QOS AF et BE.

Le coût supplémentaire induit par le calcul de la solution optimale semble justifié par les performances supérieures apportées. De plus, l'utilisation du DRR permet un passage simplifié entre la solution optimale continue et la solution effective, discrète.

On peut néanmoins remarquer les mauvaises performances, en termes de temps d'attente, des trafics AF et BE. Celles-ci ne sont pas seulement liées à l'algorithme lui-même, mais au profil très particulier du trafic (en rafales ou *burst*), qui est complexe à prendre en compte dans un système de type DAMA. De plus, le volume relativement faible de ce trafic le rend de fait moins prioritaire que les requêtes engendrées par EF. Une étude plus poussée de l'influence du trafic, et surtout de la formation des requêtes, serait donc ici un ajout significatif.

4.4 Conclusion sur l'allocation de ressources voie Retour

Dans ce chapitre, nous avons présenté une modélisation du problème d'allocation de ressources pour la voie Retour satellite, utilisant le protocole DVB-RCS2. Bien que la performance du système dépende de très nombreux paramètres, nous avons pu isoler le processus d'allocation de ressources afin de montrer son influence sur cette performance.

Notre objectif, qui consistait à proposer une solution qui, à partir d'un ensemble de requêtes, puisse donner une allocation, a été rempli par deux algorithmes, l'un *linéaire* et l'autre *optimal*. Cette solution devait également prendre en compte de manière intégrée les cas de sous-allocation (FCA) et de congestion, répartissant au mieux, dans les deux cas, les ressources disponibles.

Si l'algorithme linéaire permet de répondre à ce cahier des charges, il s'apparente plus à une solution de type priorité stricte, privilégiant la QOS la plus élevée au détriment des autres, et ne permet pas de prendre en compte l'efficacité spectrale, risquant ainsi de donner des résultats inéquitables. L'intérêt de notre proposition, reposant sur les fonctions d'utilité, est d'exprimer plusieurs comportements en une seule formulation : par l'intermédiaire des paramètres α et β , et de la fonction u définie, nous pouvons exprimer la relation entre satisfaction de la requête, contraintes de QOS, et efficacité spectrale.

Nous avons adopté une formulation sous forme d'un problème d'optimisation, résolu pour chaque Supertrame, et donnant l'allocation optimale. Nous avons pu observer que cet algorithme présentait de meilleurs résultats que l'algorithme linéaire de référence, particulièrement pour les trafics peu contraints.

Nous avons cependant noté que la seule étude de l'allocation de ressources n'était

pas suffisante pour permettre une bonne performance du système dans son ensemble : l'estimation des requêtes de trafic, ainsi que le type de requêtes (RBDC ou VBDC) ont une influence significative. De plus, nous avons délibérément laissé hors de l'algorithme la prise en compte de la variabilité de la capacité et de la demande, puisque celui-ci est «instantanée» : il donnera un résultat identique pour un même ensemble de requêtes. Ce comportement pourrait être amélioré par une prise en compte des allocations passées, mais l'introduction d'une mémoire supplémentaire dans un système déjà très complexe pourrait être préjudiciable à notre compréhension du fonctionnement de ce même système.

CONCLUSION

Le besoin d'accéder à Internet partout et à chaque instant a poussé les télécommunications par satellite sur le devant de la scène, car elles constituent une réponse bien adaptée pour compléter les réseaux terrestres, notamment dans les zones mal desservies ou non couvertes. Si de nombreux projets de constellations ont été annoncés très récemment, les solutions reposant sur un satellite en orbite géostationnaire conservent un intérêt indéniable, du fait de leur très large couverture, de leur capacité importante, toutes deux obtenues avec un seul satellite. Les standards DVB-S2/GSE et DVB-RCS2 sont maintenant matures pour les applications d'accès à Internet par satellite (*broadband*). Ils permettent d'apporter aux utilisateurs des services comparables aux réseaux terrestres : téléphone, vidéo et accès à Internet, utilisant une architecture IP unifiée.

Cependant, le passage aux nouvelles générations de standards, DVB-S2 et DVB-RCS2, constitue une évolution majeure : si les systèmes de diffusion de télévision ne présentaient pas de difficulté particulière quant à l'ordonnancement, il faut à présent composer avec un débit variable, inhérent aux applications du monde IP, et une capacité elle aussi variable, du fait de l'utilisation de méthode d'adaptation de type modulation et codage adaptatif (ACM), sans oublier le temps de propagation du satellite, particulièrement long. Nous avons traduit cette variabilité en considérant deux processus évoluant de manière décorrélée : la demande, provenant des utilisateurs du système, et la capacité, tributaire des conditions météorologiques. Cette variabilité a un impact considérable sur l'ordonnancement et l'allocation de ressources, puisqu'il faut s'adapter en permanence à l'évolution de la demande et de la capacité, tout en assurant un niveau de service satisfaisant, autrement dit des garanties de QoS. Ce problème, notoirement complexe, a suscité de très nombreuses recherches, dans le terrestre comme dans le satellite.

Nous avons, à travers notre étude des solutions proposées dans la littérature, montré comment les solutions actuelles, outre le fait que beaucoup ne reposent pas sur des architectures de dernière génération, ne sont pas adaptées à la prise en compte conjointe des problèmes que nous avons exposés. Ceux-ci concernent la variabilité de la demande,

de la capacité, contraintes de QoS, et les exigences de complexité pour l'implantation en temps réel. Nous avons également remarqué que beaucoup d'algorithmes sont pénalisés par leur manque d'anticipation : ils ne permettent pas d'évaluer l'impact des décisions d'ordonnancement sur le système. Ainsi, une solution pertinente d'ordonnancement ou d'allocation de ressources doit prendre en compte la dynamique du système, ainsi que sa complexité, et son impact sur l'évolution du système.

Sur la voie Aller, nous avons retenu deux mécanismes : les règles d'ordonnancement, pour leur simplicité de mise en œuvre et l'apport de très nombreux travaux de recherche, et les fonctions d'utilité, pour leur flexibilité et leur anticipation à long terme. Nous avons montré comment adapter les règles d'ordonnancement, d'abord conçues pour le monde terrestre, au satellite et à ses contraintes. Concernant les fonctions d'utilité, nous avons modélisé l'ordonnancement comme solution, à long terme, d'un problème d'optimisation. Nous avons montré comment il était possible d'approcher la solution de ce problème à l'aide d'un mécanisme d'optimisation itératif, dont le système effectue une étape à chaque trame de la couche physique, ou BBFRAME, et proposé deux algorithmes, à l'échelle d'une (BBS) ou plusieurs trames (BPS). Les résultats de simulations sur plusieurs scénarios montrent l'efficacité de BPS, qui constitue la solution la plus adaptée à notre problème, possédant à la fois une vision à long terme de l'ordonnancement, une implantation simple, et une prise en compte explicite de la QoS. Cet algorithme semble être mature pour une implantation opérationnelle.

Sur la voie Retour, nous avons proposé deux algorithmes, l'un *linéaire*, l'autre *optimal*, tous deux optimisant l'allocation Supertrame par Supertrame. Nous avons pu noter que ce problème était bien plus complexe que pour la voie Aller, car dépendant de nombreux processus, tels que l'estimation de requêtes de trafic dans DVB-RCS2, dont l'influence est importante. Notre algorithme optimal repose sur une modélisation, à l'aide de fonctions d'utilité, de la valorisation d'une allocation en fonction de la requête, des contraintes de QoS, et de l'efficacité spectrale. Ainsi, notre allocation de ressources peut s'exprimer sous forme d'un autre problème d'optimisation, pour lequel nous avons proposé une résolution utilisant le problème dual équivalent. Nous avons pu observer, à travers nos simulations, que notre algorithme optimal permettait une bien meilleure prise en charge des QoS peu prioritaires en comparaison de l'algorithme linéaire, tout en assurant un niveau de service très satisfaisant pour les plus prioritaires.

Nous disposons à présent d'une réponse satisfaisante au problème posé au début de notre étude, c'est-à-dire d'algorithmes d'ordonnancement et d'allocation de ressource, performants et peu complexes, bien adaptés au contexte, et particulièrement à celui des standards satellite. Cependant, nous avons pu remarquer, tout au long de notre étude, que

de nombreux points restent à éclaircir, ou à développer, afin d'affiner notre compréhension des nombreux processus régissant les performances des Systèmes de Télécommunications par Satellite :

- Sur la voie Aller, il serait intéressant d'introduire des modèles de trafic plus détaillés, comme par exemple en exploitant une source de trafic pour chaque Terminal. Cette précision permettrait non seulement d'observer le comportement de notre algorithme vis-à-vis d'un trafic plus réaliste, mais surtout nous serions à même de mener une étude sur le multiplexage des flux, ici fait au niveau IP, qui peut avoir une influence importante sur la performance de bout-en-bout.
- Nous avons majoritairement considéré une distribution de MODCOD fixe, or celle-ci est susceptible de changer, suivant les conditions météorologiques au-dessus du spot du satellite. Des événements de forte pluie peuvent potentiellement concerner une part non négligeable des Terminaux et perturber le fonctionnement du système entier, puisque leurs besoins en ressources vont très rapidement augmenter. Ainsi, une étude de ces événements météorologiques, de leur durée, période et intensité, permettrait de se prémunir de situations potentiellement instables, même si des prédictions sur ce type d'échelles de temps restent délicates.
- Notre évaluation de performances s'est faite selon des critères de qualité de service, tels que le délai et la gigue. Cependant, une évaluation de la qualité de l'expérience de l'utilisateur serait également pertinente dans notre cas, étant donné que nous considérons un système fournissant des services diversifiés (voix, vidéo, Internet). Ainsi, la relation entre ordonnancement et qualité de la vidéo, de la conversation téléphonique, ou bien le temps de réponse moyen d'une page web pourraient nous éclairer sur les possibilités d'amélioration de nos algorithmes.
- Nous n'avons considéré ici qu'un schéma d'allocation type DAMA, mais il est également possible d'utiliser une méthode d'accès aléatoire de type CRDSA, bien adaptée au trafic asynchrone de type web. Il serait alors intéressant d'analyser les performances conjointes de ces deux méthodes, notamment en considérant une trame DVB-RCS2 où la proportion de chaque BTU par méthode pourrait s'adapter dynamiquement au trafic.
- De nombreux travaux de recherche préconisent la réutilisation dynamique de l'ensemble de la bande passante sous la couverture. Ce paradigme engendrerait des gains considérables en capacité, car il nous offrirait une dimension supplémentaire dans notre allocation de ressources, mais au prix de calculs d'interférences très complexes. Les outils que nous avons développés ici semblent adaptés à un tel problème, mais ils nécessiteraient une adaptation de la formulation pour intégrer la

contrainte de niveau d'interférences, et une dimension supplémentaire, la porteuse choisie.

- Enfin, nous avons pu remarquer la très grande sensibilité des algorithmes à l'évaluation de performances utilisant des simulations. Aussi, les outils et méthodes que nous avons développé dans cette thèse offrent avant tout les outils nécessaires à la formalisation et à l'évaluation des problèmes d'ordonnancement et d'allocation de ressources. Les fonctions d'utilité dont nous avons fait usage restent très simples, particulièrement pour la voie Retour, et il serait certainement bénéfique de proposer une relation plus réaliste entre qualité de service, efficacité spectrale, et requête de trafic correspondante.

PUBLICATIONS

- J.-B. Dupé, E. Chaput, C. Baudoin, C. Bes, A. Deramecourt, and A.-L. Beylot, “Rule-based Packet Scheduling for DVB-S2 Through Generic Stream Encapsulation,” in *IEEE International Conference on Communications (ICC)*, pp. 3582–3587, June 2014
- J.-B. Dupé, E. Chaput, C. Baudoin, C. Bes, A. Deramecourt, and A.-L. Beylot, “Optimized gse packet scheduling over dvb-s2,” in *IEEE Global Communications Conference (GLOBECOM)*, pp. 2856–2861, Dec 2014

BIBLIOGRAPHIE

- [1] EBU-UER, “Digital video broadcasting (dvb); framing structure, channel coding and modulation for 11/12 ghz satellite services,” Tech. Rep. EN 300 421 v1.1.2, European Telecommunications Standards Institute, August 1997.
- [2] EBU-UER, “Digital video broadcasting (dvb); interaction channel for satellite distribution systems,” Tech. Rep. EN 301 790 v1.3.1, European Telecommunications Standards Institute, March 2003.
- [3] EBU-UER, “Digital video broadcasting (DVB); second generation framing structure, channel coding and modulation systems for broadcasting, interactive services, news gathering and other broadband satellite applications,” Tech. Rep. v1.1.2, European Telecommunications Standards Institute, June 2006.
- [4] F. Vieira and M. A. Vazquez-Castro, “A tunable-fairness cross-layer scheduler for DVB-S2,” *International Journal of Satellite Communications and Networking*, vol. 24, pp. 437–450, 2006.
- [5] G. Fairhurst, G. Giambene, S. Giannetti, C. Parraga, and A. Sali, “Multimedia traffic scheduling in DVB-S2 networks with mobile users,” *2008 IEEE International Workshop on Satellite and Space Communications*, pp. 211–215, Oct. 2008.
- [6] E. Chaput, A.-L. Beylot, and C. Baudoin, “Packet Scheduling Over DVB-S2 Through GSE Encapsulation,” *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–5, 2008.
- [7] D. Pradas and M. A. Vazquez-Castro, “NUM-Based Fair Rate-Delay Balancing for Layered Video Multicasting over Adaptive Satellite Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 969–978, May 2011.
- [8] E. Rendón-Morales, J. Mata-Díaz, J. Alins, J. L. Muñoz, and O. Esparza, “QoSAr : a cross-layer architecture for E2E QoS provisioning over DVB-S2 broadband satellite systems,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 302, 2012.
- [9] J.-B. Dupé, E. Chaput, C. Baudoin, C. Bes, A. Deramecourt, and A.-L. Beylot, “Rule-based Packet Scheduling for DVB-S2 Through Generic Stream Encapsulation,” in *IEEE International Conference on Communications (ICC)*, pp. 3582–3587, June 2014.

- [10] EBU-UER, “Digital video broadcasting (DVB) user guidelines for the second generation system for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2),” Tech. Rep. v1.1.1, European Telecommunications Standards Institute, February 2005.
- [11] J. V. Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [12] DVB-GBS. W. Group, “IP/s.2 study, GSE protocol specification,” Tech. Rep. Document rev 10, European Telecommunications Standards Institute, 2007.
- [13] D. Pradas, *Cross-layer design and methodology for satellite broadband networking (PhD Thesis)*. Sept 2011.
- [14] C. Berrou and A. Glavieux, “Near optimum error correcting coding and decoding : turbo-codes,” *IEEE Transactions on Communications*, vol. 44, pp. 1261–1271, Oct 1996.
- [15] D. J. MacKay and R. M. Neal, “Near shannon limit performance of low density parity check codes,” *Electronics Letters*, vol. 32, pp. 1645–1646, 1996.
- [16] ETSI, “Digital video broadcasting (dvb) ; second generation dvb interactive satellite system (dvb-rs2) ; part 2 : Lower layers for satellite standard,” Tech. Rep. V1.1.1, European Telecommunications Standards Institut, August 2011.
- [17] E. Casini, R. De Gaudenzi, and O. Herrero, “Contention resolution diversity slotted aloha (crdsa) : An enhanced random access scheme for satellite access packet networks,” *IEEE Transactions on Wireless Communications*, vol. 6, pp. 1408–1419, April 2007.
- [18] S. Cioni, R. D. Gaudenzi, and R. Rinaldo, “Channel estimation and physical layer adaptation techniques for satellite networks exploiting adaptive coding and modulation,” *International Journal of Satellite Communications and Networking (IJSCN)*, pp. 157–188, 2008.
- [19] L. Chisci, R. Fantacci, and T. Pecorella, “Predictive bandwidth control for geo satellite networks,” in *IEEE International Conference on Communications (ICC)*, vol. 7, pp. 3958–3962 Vol.7, June 2004.
- [20] Eutelsat, “Eutelsat ka-sat.” <http://www.eutelsat.com/fr/satellites/flotte/EUTELSAT-KA-SAT.html>.
- [21] R. Sinha, C. Papadopoulos, and J. Heidemann, “Internet packet size distributions : Some observations,” May 2007. Originally released October 2005 as web page <http://netweb.usc.edu/~rsinha/pkt-sizes/>.
- [22] G. Maral and M. Bousquet, *Satellite communications systems : systems, techniques and technology*. Wiley series in communication and distributed systems, Wiley, 1998.
- [23] N. Jeannin, X. Boulanger, L. Féral, L. Castanet, and F. Lacoste, “Inter-annual variability, risk and confidence intervals associated with propagation statistics. part I : theory of estimation,” *International Journal on Satellite Communications Networking (IJSCN)*, vol. 32, no. 5, pp. 407–421, 2014.

- [24] J.-B. Dupé, E. Chaput, C. Baudoin, C. Bes, A. Deramecourt, and A.-L. Beylot, "Optimized gse packet scheduling over dvb-s2," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 2856–2861, Dec 2014.
- [25] G. Song, Y. Li, L. Cimini, and H. Zheng, "Joint channel-aware and queue-aware data scheduling in multiple shared wireless channels," *2004 IEEE Wireless Communications and Networking Conference (IEEE WCNC)*, pp. 1939–1944, 2004.
- [26] P. Hosein, "QoS control for WCDMA high speed packet data," *4th International Workshop on Mobile and Wireless Communications Network*, no. 4, pp. 169–173, 2002.
- [27] S. Shakkottai and A. L. Stolyar, "Scheduling algorithms for a mixture of real-time and non-real-time data in HDR," *17th International Teletraffic Congress (ITC)*, pp. 793–804, 2000.
- [28] E. Chaput, M. Verloop, A.-L. Beylot, and C. Baudoin, "Utility function based packet scheduling over DVB-S2," in *2013 IEEE International Conference on Communications (ICC)*, pp. 4288–4292, June 2013.
- [29] X. Wang and G. B. Giannakis, "Resource Allocation for Wireless Multiuser OFDM Networks," *IEEE Transactions on Information Theory*, vol. 57, pp. 4359–4372, July 2011.
- [30] S. J. Golestani and S. Street, "A Self-Clocked Fair Queueing Scheme for Broadband Applications," pp. 636–646, 1994.
- [31] A. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," *SIGCOMM Comput. Commun. Rev.*, vol. 19, pp. 1–12, Aug. 1989.
- [32] J. Bennett and H. Zhang, "Wf2q : worst-case fair weighted fair queueing," vol. 1, pp. 120–128 vol.1, Mar 1996.
- [33] E. Rendon-Morales, J. Mata-Díaz, J. Alins, J. L. Muñoz, and O. Esparza, "Cross-layer packet scheduler for QoS support over Digital Video Broadcasting-Second Generation broadband satellite systems," *International Journal of Communication Systems*, Oct. 2012.
- [34] M. Tropea, F. Veltri, F. De Rango, A.-F. Santamaria, and L. Belcastro, "Two Step Based QoS Scheduler for DVB-S2 Satellite System," *2011 IEEE International Conference on Communications (ICC)*, pp. 1–5, June 2011.
- [35] E. Hahne, "Round-robin scheduling for max-min fairness in data networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 1024–1039, Sep 1991.
- [36] H. Shimonishi, M. Yoshida, R. Fan, and H. Suzuki, "An improvement of weighted round robin cell scheduling in atm networks," in *IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, pp. 1119–1123 vol.2, Nov 1997.
- [37] M. Katevenis, S. Sidiropoulos, and C. Courcoubetis, "Weighted round-robin cell multiplexing in a general-purpose atm switch chip," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 1265–1279, Oct 1991.

- [38] M. Shreedhar and G. Varghese, "Efficient fair queuing using deficit round-robin," *IEEE/ACM Transactions on Networking*, vol. 4, pp. 375–385, June 1996.
- [39] J. Nagle, "On Packet Switches with Infinite Storage," *IEEE Transactions on Communications*, vol. 35, no. 4, pp. 435–438, 1987.
- [40] J. Bennett and H. Zhang, "Wf2q : worst-case fair weighted fair queueing," *IEEE International Conference on Computer Communications (INFOCOM)*, vol. 1, pp. 120–128 vol.1, Mar 1996.
- [41] J. Bennett, "Hierarchical packet fair queueing algorithms," *IEEE/ACM Transactions on Networking*, vol. 5, no. 5, pp. 675–689, 1997.
- [42] X. Yuan and Z. Duan, "Frr : a proportional and worst-case fair round robin scheduler," *IEEE International Conference on Computer Communications (INFOCOM)*, vol. 2, pp. 831–842 vol. 2, March 2005.
- [43] X. Fei and Marshall, "DO-WF2Q : delay-optimised WF2Q packet scheduling," *IEEE International Conference on Communications (ICC)*, vol. 2, pp. 1080–1085, 2002.
- [44] M. Karsten, "Si-wf2q : Wf2q approximation with small constant execution overhead," *IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1–12, April 2006.
- [45] U. Ayesta, M. Erausquin, M. Jonckheere, and I. Verloop, "Scheduling in a random environment : Stability and asymptotic optimality," *IEEE/ACM Transactions on Networking*, vol. 21, pp. 258–271, Feb 2013.
- [46] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, pp. 33–37, 1997.
- [47] S. Shakkottai and A. L. Stolyar, "Scheduling for Multiple Flows Sharing a Time-Varying Channel : The Exponential Rule," *Translations of the American Mathematical Society-Series 2*, vol. 207, pp. 185—202, 2002.
- [48] F. Capozzi, G. Piro, L. Grieco, G. Boggia, and P. Camarda, "Downlink packet scheduling in lte cellular networks : Key design issues and a survey," *IEEE Communications Surveys Tutorials*, vol. 15, pp. 678–700, Second 2013.
- [49] D. Ferrari and D. Verma, "A scheme for real-time channel establishment in wide-area networks," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 8, pp. 368–379, Apr 1990.
- [50] J. Nash, John F., "The bargaining problem," *Econometrica*, vol. 18, no. 2, pp. pp. 155–162, 1950.
- [51] E. Altman, T. Boulogne, R. El-Azouzi, T. Jiménez, and L. Wynter, "A survey on networking games in telecommunications," *Computers & Operations Research*, vol. 33, pp. 286–311, Feb. 2006.
- [52] H. Yaiche, R. Mazumdar, and C. Rosenberg, "A game theoretic framework for bandwidth allocation and pricing in broadband networks," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667–678, 2000.

- [53] Z. Han, Z. Ji, and K. Liu, "Fair Multiuser Channel Allocation for OFDMA Networks Using Nash Bargaining Solutions and Coalitions," *IEEE Transactions on Communications*, vol. 53, pp. 1366–1376, Aug. 2005.
- [54] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks : Shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, no. 3, pp. pp. 237–252, 1998.
- [55] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 226–234, Feb. 2009.
- [56] G. Song and Y. Li, "Utility-based resource allocation and scheduling in ofdm-based wireless broadband networks," *IEEE Communications Magazine*, vol. 43, pp. 127–134, Dec 2005.
- [57] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277–1294, Jun 2002.
- [58] G. Song, Y. Li, and L. J. Cimini, "Joint channel- and queue-aware scheduling for multiuser diversity in wireless OFDMA networks," *IEEE Transactions on Communications*, vol. 57, no. 7, pp. 2109–2121, 2009.
- [59] N. Zhou, X. Zhu, Y. Huang, and H. Lin, "Low complexity cross-layer design with packet dependent scheduling for heterogeneous traffic in multiuser OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 1912–1923, June 2010.
- [60] I.-H. Hou and P. R. Kumar, "Utility Maximization for Delay Constrained QoS in Wireless," pp. 1–9, Mar. 2010.
- [61] G. Fairhurst, G. Giambene, S. Giannetti, C. Niebla, and A. Sali, "Cross-layer study for resource management in dVB-S2 with mobile users," in *International Workshop on Satellite and Space Communications (IWSSC)*, pp. 257–261, Sept 2009.
- [62] M. Castro and G. Granados, "Cross-layer packet scheduler design of a multibeam broadband satellite system with adaptive coding and modulation," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 248–258, Jan 2007.
- [63] G. Seco-Granados, M. A. Vazquez-Castro, A. Morell, and F. Vieira, "SAT03-6 : Algorithm for Fair Bandwidth Allocation with QoS Constraints in DVB-S2/RCS," in *IEEE Global Communications Conference (GLOBECOM)*, pp. 1–5, Nov. 2006.
- [64] M. A. Vazquez-Castro and F. Vieira, "DVB-S2 Full Cross-Layer Design for QoS Provision," *IEEE Communications Magazine*, no. January, pp. 128–135, 2012.
- [65] S. Martello and P. Toth, *Knapsack Problems : Algorithms and Computer Implementations*. New York, NY, USA : John Wiley & Sons, Inc., 1990.
- [66] D. Pradas, A. Bouabdallah, J. Lacan, M. Vazquez Castro, and M. Bousquet, "Cross-layer optimization of unequal protected layered video over hierarchical modulation," in *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1–6, Nov 2009.

- [67] H. Radha, M. van der Schaar, and Y. Chen, "The mpeg-4 fine-grained scalable video coding method for multimedia streaming over ip," *IEEE Transactions on Multimedia*, vol. 3, pp. 53–68, Mar 2001.
- [68] K. Nichols, S. Blake, F. Baker, and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers." RFC 2474 (Proposed Standard), Dec. 1998. Updated by RFCs 3168, 3260.
- [69] P. Hande, Z. Shengyu, and M. Chiang, "Distributed rate allocation for inelastic flows," *IEEE/ACM Transactions on Networking*, vol. 15, pp. 1240–1253, Dec 2007.
- [70] A. Beylot, R. Dhaou, and C. Baudoin, "Resource allocation in DVB-RCS satellite systems," in *IFIP-TC6 International Networking Conference (Networking)*, pp. 511–521, 2007.
- [71] B. de la Cuesta, L. Albiol, J. Aguiar, C. Baladron, B. Carro, and A. Sanchez-Esguevillas, "Innovative dama algorithm for multimedia dvb-rcs system," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, p. 14, 2013.
- [72] N. Celandroni, F. Davoli, and E. Ferro, "Static and dynamic resource allocation in a multiservice satellite network with fading," *International Journal of Satellite Communications and Networking*, vol. 21, no. 4-5, pp. 469–487, 2003.
- [73] A. Pietrabissa and A. Fiaschetti, "Dynamic uplink frame optimization with adaptive coding and modulation in DVB-RCS2 satellite networks," *International Journal on Satellite Communications and Networking (IJSCN)*, no. March, pp. 123–139, 2013.
- [74] J. Chu, N. Dukkipati, Y. Cheng, and M. Mathis, "Increasing TCP's Initial Window." RFC 6928 (Experimental), Apr. 2013.
- [75] E. Del Re, G. Gorni, L. Ronga, and M. Castro, "A game theory approach for dvb-rs resource allocation," in *IEEE Vehicular Technology Conference (VTC Spring)*, pp. 2937–2941, May 2008.
- [76] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2008.
- [77] J. Gettys, "Bufferbloat : Dark Buffers in the Internet," *IEEE Internet Computing*, vol. 15, pp. 96–96, May 2011.
- [78] S. Stefanov, "Convex separable minimization subject to bounded variables," *Computational Optimization and Applications*, no. 2, pp. 27–48, 2001.
- [79] E. Chaput, "Ndes, network discrete event simulator." <https://github.com/Manu-31/ndes>.
- [80] B. Davie, A. Charny, J. Bennet, K. Benson, J. L. Boudec, W. Courtney, S. Davari, V. Firoiu, and D. Stiliadis, "An Expedited Forwarding PHB (Per-Hop Behavior)." RFC 3246 (Proposed Standard), Mar. 2002.
- [81] G. Giambene, *Resource Management in Satellite Networks*. Springer, 2007.

-
- [82] F. Delli Priscoli and A. Pietrabissa, “Control-theoretic bandwidth-on-demand protocol for satellite networks,” in *IEEE International Conference on Control Applications*, vol. 1, pp. 530–535 vol.1, 2002.
- [83] 3GPP2, “cdma2000 evaluation methodology, revision a,” Tech. Rep. C.R1002-A, May 2009.

Résumé

Les télécommunications par satellite ont connu ces dernières années un regain d'intérêt significatif, du fait de leur capacité à réduire la fracture numérique. Le problème de l'ordonnancement dans ces systèmes est apparu comme particulièrement important, puisqu'il doit prendre en compte deux processus évoluant de manière totalement décorrélée : la demande des utilisateurs et la capacité du système. Cette thèse s'intéresse au problème de l'ordonnancement et de l'allocation de ressources, dans le but de fournir un service comparable, complémentaire aux réseaux terrestres. Nous présentons ainsi deux algorithmes, l'un pour la voie Aller, l'autre pour la voie Retour, qui permettent un très bon compromis entre demande et capacité. Nous avons de plus montré que ces algorithmes étaient matures pour une utilisation dans un système réel.

Mots clés : Ordonnancement, DVB-S2, DVB-RCS2, Allocation de ressources, QoS, Optimisation

Abstract

Satellite telecommunications have seen a tremendous increase in interest, due to its ability to reduce the digital divide. Scheduling has arisen as a major issue of those modern systems, since it has to deal with to highly uncorrelated process : demand and capacity. This thesis aims to study the problem of scheduling and resource allocation in satellite systems, hoping to achieve a service that can match with terrestrial networks, and focused on remote areas. We thus present two algorithm developed : one for the Forward link, one for the Return link, who both allow a very good trade-off between demand and capacity. We have furthermore shown that our algorithms were mature for industrial implementation.

Keywords : Scheduling, DVB-S2, DVB-RCS2, Resource Allocation, QoS, Optimization