



**HAL**  
open science

# Asymptotically Optimal Scheduling Schemes for Large Networks

Saad Kriouile

► **To cite this version:**

Saad Kriouile. Asymptotically Optimal Scheduling Schemes for Large Networks. Networking and Internet Architecture [cs.NI]. Université Paris-Saclay, 2021. English. NNT : 2021UPASG022 . tel-03243493

**HAL Id: tel-03243493**

**<https://theses.hal.science/tel-03243493>**

Submitted on 31 May 2021

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

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

# Asymptotically Optimal Scheduling Schemes for Large Networks

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de l'Information  
et de la Communication (STIC)

Spécialité de doctorat: réseaux, information et communications

Unité de recherche: Université Paris-Saclay, CNRS, CentraleSupélec,  
Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France.

Référent: CentraleSupélec, 91190 Gif sur Yvette, France

Thèse présentée et soutenue à Paris Saclay, le 25/03/2021, par

**Saad KRIOULE**

## Composition du Jury:

<b>Inbar Fijalkow</b> Professeure, ENSEA	Présidente
<b>Nikolaos Pappas</b> Associate Professor, Linkoping University	Rapporteur & examinateur
<b>Rachid El Azouzi</b> Professeur, Université d'avignon	Rapporteur & examinateur
<b>Omur Ozel</b> Assistant Professor, George Washington University	Examineur
<b>Samson Lasaulce</b> DR, CNRS	Examineur

## Direction de la thèse:

<b>Mohamad Assaad</b> Professeur, CentraleSupélec	Directeur
--	-----------

**Titre:** Les schémas d'ordonnement asymptotiquement optimaux pour les grands réseaux

**Mots clés:** allocation des ressources, Temps d'attente, Age de l'information, indice de Whittle, Relaxation Lagrangian, optimalité asymptotique

**Résumé:** Au courant de la dernière décennie, le nombre de machines connectées aux réseaux sans fils, a connu une croissance exponentielle à cause de l'apparition de l'Internet des Objets. Ainsi, le réseau sans fils 5G, est apparu pour accompagner cette grande évolution technologique. Parmi les exigences de la 5G est d'optimiser certaines métriques qui permettent une communication fiable et fluide, à savoir la latence des paquets et l'âge de l'information. Dans ce cadre, un problème majeur consiste à savoir comment partager les ressources disponibles entre le nombre massif des utilisateurs dans le réseau de telle façon à optimiser ces métriques en question. Pour répondre à cette problématique, nous étudions dans cette thèse le problème d'allocation des canaux où le nombre d'utilisateurs dans le système est supérieur à celui des canaux. Notre but est de trouver une solution qui permet d'allouer les canaux disponibles à un sous-ensemble d'utilisateurs à chaque instant, de manière à minimiser l'espérance de la moyenne à long terme du temps d'attente des paquets avant la transmission ou bien l'espérance de la moyenne à long terme de l'âge de l'information. Puisque le problème étudié entre dans le cadre des problèmes de Restless bandits, la solution optimale est hors de portée. Néanmoins, pour contourner cette difficulté, nous adoptons une approche basée sur les indices de Whittle pour obtenir une méthode d'allocation appelée "Whittle Index Policy" (WIP) qui est très performante avec une complexité faible, surtout lorsque le nombre d'utilisateurs et des canaux est suffisamment grand. Dans cette mesure, nous expliquons en détail, au chapitre 2, l'approche pour obtenir les indices de Whittle. Ensuite, nous étudions au chapitre 3 le problème d'allocation des canaux dans un système de files d'attente avec pour objectif de minimiser le temps d'attente des paquets dans les files. Nous

adoptons une méthode basée sur la Relaxation Lagrangienne et nous démontrons que la solution optimale de l'équation de Bellman du problème dual est de type solution à seuils (threshold-based solution). Ensuite, nous prouvons que ce problème est indexable trouvons par la suite l'expression des indices de Whittle, en distinguant si la taille maximale des files est finie ou infinie. Puis, nous prouvons par une démonstration mathématique rigoureuse que notre solution est effectivement optimale dans le cas où le nombre d'utilisateurs est infiniment grand. Enfin, nous donnons des résultats numériques qui mettent en évidence la bonne performance de notre solution proposée et qui confirment nos résultats théoriques. Dans le chapitre 4, nous examinons le problème d'allocation des canaux dans un réseau sans fil avec pour objectif de minimiser l'âge de l'information moyen. Comme le problème étudié dans le chapitre 3 et de façon similaire, trouver une solution optimale n'est pas évident. De ce fait, nous adoptons également l'approche basée sur les indices de Whittle pour développer une politique d'allocation simple et performante. Notre principale contribution dans ce chapitre consiste d'une part à donner des résultats rigoureux sur l'optimalité asymptotique de la politique de Whittle (WIP) dans un régime où le nombre d'utilisateurs et des canaux tend vers l'infini. En effet, nous présentons une nouvelle approche mathématique pour établir l'optimalité lorsque l'âge de l'information n'est pas borné par une certaine valeur. Cette nouvelle approche est basée sur des techniques complexes (critères de Cauchy, etc.), et contrairement aux travaux précédents, la démonstration n'exige aucune hypothèse simplificatrice sur le système considéré. Finalement, nous présentons des résultats numériques qui montrent l'optimalité de la politique de Whittle et qui confirment nos résultats théoriques.

**Title:** Asymptotically Optimal Scheduling Schemes for Large Networks

**Keywords:** resources allocation, queuing delay, Age of information, Whittle's indices, Lagrangian Relaxation, Asymptotic optimality

**Abstract:** In this thesis, we investigate a general channel allocation problem where the number of channels is less than that of users. The aim is to find a policy that schedules the channels to a given subset of users at each time slot in such a way to minimize two different objectives functions namely, the long-run expected average queuing delay (chapter 3) and the long-run expected average age of information (Chapter 4). We show that our problems fall in the framework of Restless Bandit Problems (RBP), for which obtaining the optimal solution is known to be out of reach. To circumvent this difficulty, we tackle the problem by adopting a Whittle index approach. In Chapter 2, we explain the Lagrangian relaxation, the steady-state and discounted cost approaches used to obtain the expressions of Whittle indices. The structure of each subproblem's optimal solution is provided in chapter 3 and 4 depending on the system models and the considered metrics (queue length or Age of Information).

In Chapter 3, the objective of the scheduling problem is to minimize the total average backlog queues of the network in question. We apply the Lagrangian relaxation approach detailed in Chapter 2 for the present model, and we prove that the optimal solution of the one-dimensional problem is of type threshold policy. After that, we establish that the aforementioned problem is indexable. Armed

with that, we apply the discounted cost approach when the queue size is infinite and the steady-state approach when the queue size is tight to obtain the Whittle indices expressions. We then provide rigorous mathematical proof that our policy is optimal in the infinitely many users regime. Finally, we provide numerical results that showcase the remarkable good performance of our proposed policy and that corroborate the theoretical findings.

In Chapter 4, we examine the average age minimization problem where users transmit over unreliable channels. Similarly to the problem studied in Chapter 3, finding the optimal scheduling scheme is known to be challenging. Accordingly, we adopt the Whittle index approach to derive the Whittle indices. Our main contribution is to provide rigorous results on the asymptotic optimality of Whittle Index Policy (WIP) in the many-users regime when the state space of the age of information is finite. However, when the state space of the Age of Information is infinite, we provide a new mathematical approach to establish the optimality of WIP for specific network settings. This novel approach is based on intricate techniques, and unlike previous works in the literature, it is free of any mathematical assumptions. Finally, we lay out numerical results that corroborate our theoretical findings and demonstrate the policy's notable performance in the many-users regime.





---

# Acknowledgements

First and foremost, I would like to thank **God** for all his blessings and for providing me the opportunity to work in a research field that I appreciate.

I would like to express my sincere gratitude to my advisor Prof. **Mohamad Assaad** for the continuous support of my Ph.D. study and research, for his motivation, for his encouragement during my thesis. He offered me generously his time and gave me relevant advice about technical and non-technical issues regarding my thesis and its development. He always had faith in me and encouraged me to move forward with my research and my studies. His guidance helped me in all the time of research and writing of this thesis.

I also express my deepest thanks to Doctor **Larranaga Maialen** for giving me the necessary advice and guidance, which were extremely valuable for my progress concerning my work during my internship, which is an initiation of my thesis.

I would also like to express my appreciation to the jury committee members: **Nikkos Pappas**, **Rachid El Azouzi**, **Inbar Fijalkow**, **Omur Ozel**, and **Samson Lasaulce** for the time spent reading this work and for all their valuable feedback on it.

Thanks also go out to my office colleague **Ali Maatouk** with whom I try to solve some issues confronted in our collaborative works. I thank him also for his valuable assistance in writing articles.

I would also like to thank the professors who helped me through skills training to improve my professional attitude. I thank all my friends who support me during my time in French. I would also like to express my gratitude to my brother **Yassine** for his constant support during my thesis, and all my brothers and sisters who always encouraged me. I am extremely grateful to my beloved parents in Morocco for their unlimited support during my years in France and all my life. Thank you for your sacrifices, your love, your encouragement, and your trust. Special thanks to my mother for her crucial role and invaluable assistance. I am forever indebted to her.



# Contents

<b>1</b>	<b>Introduction en français</b>	<b>7</b>
1.1	Motivation . . . . .	7
1.2	Ordonnancement basé sur le temps d'attente . . . . .	8
1.3	Ordonnancement basé sur l'âge . . . . .	9
1.4	Etat de l'Art . . . . .	13
1.4.1	Problème d'ordonnancement: Minimisation du temps d'attente . . .	13
1.4.2	Problème d'ordonnancement: Minimisation de l'Age . . . . .	15
1.5	Contributions et Aperçu de la thèse . . . . .	16
<b>2</b>	<b>Introduction</b>	<b>19</b>
2.1	Motivation . . . . .	19
2.2	Delay-based Scheduling . . . . .	20
2.3	Age-based Scheduling . . . . .	22
2.4	State of the Art . . . . .	25
2.4.1	Scheduling Problem: Delay Minimization . . . . .	25
2.4.2	Scheduling Problem: Age Minimization . . . . .	26
2.5	Contributions and overview of the thesis . . . . .	28
<b>3</b>	<b>Methodology</b>	<b>29</b>
3.1	Restless Bandit Problem . . . . .	29
3.2	System model of interest . . . . .	30
3.3	Lagrangian approach and Whittle index policy . . . . .	30
3.3.1	Relaxed Problem and Dual Problem . . . . .	31
3.3.2	Problem Decomposition . . . . .	31
3.3.3	Whittle's Index . . . . .	33
3.4	Asymptotic Optimality of Whittle Index Policy . . . . .	36
3.4.1	Local Optimality . . . . .	36
3.4.2	Global Optimality . . . . .	37
<b>4</b>	<b>Whittle's index policy for minimizing the Delay in Queuing systems</b>	<b>39</b>
4.1	System Model . . . . .	39
4.2	Problem formulation . . . . .	40
4.3	Relaxed Problem and Threshold-based Policy . . . . .	41
4.3.1	Relaxed Problem and Dual Problem . . . . .	41
4.3.2	Problem Decomposition and Threshold-based Policy . . . . .	41
4.4	Whittle's Index . . . . .	47
4.4.1	Steady State approach ( $L < R$ ) . . . . .	47

4.4.2	Discounted Cost approach ( $L = +\infty$ ) . . . . .	51
4.5	Further analysis of the optimal solution of the relaxed problem . . . . .	54
4.6	Local asymptotic optimality ( $L < R$ ) . . . . .	55
4.7	Global asymptotic optimality ( $L < R$ ) . . . . .	59
4.8	Numerical Results . . . . .	60
4.8.1	Numerical Results for finite buffer size . . . . .	60
4.8.2	Numerical Results for infinite buffer size . . . . .	62
4.9	Conclusion . . . . .	64
<b>5</b>	<b>Whittle's index policy for minimizing the Age of Information</b>	<b>65</b>
5.1	System Model . . . . .	65
5.1.1	Network description . . . . .	65
5.1.2	Adopted age metric . . . . .	66
5.1.3	Problem formulation . . . . .	66
5.2	Local optimality: Finite $L$ . . . . .	70
5.3	Global Optimality: Infinite $L$ . . . . .	74
5.4	Numerical Results . . . . .	88
5.5	Conclusion . . . . .	89
<b>6</b>	<b>Conclusions and Outlook</b>	<b>91</b>
6.1	Conclusion . . . . .	91
6.2	Future Works . . . . .	91
<b>A</b>	<b>Appendix: Methodology</b>	<b>101</b>
A.1	Proof of Proposition 3.1 . . . . .	101
A.2	Proof of Proposition 3.2 . . . . .	102
<b>B</b>	<b>Appendix: Whittle's index policy for minimizing the Delay in Queuing system</b>	<b>107</b>
B.1	Proof of Lemma 4.2 . . . . .	107
B.2	Proof of Proposition 4.1 . . . . .	108
B.3	Proof of Proposition 4.2 . . . . .	109
B.4	Proof of Proposition 4.3 . . . . .	111
B.5	Proof of Theorem 4.3 . . . . .	111
B.6	Proof of Proposition 4.4 . . . . .	113
B.7	Proof of Proposition 4.5 . . . . .	116
B.8	Proof of Proposition 4.6 . . . . .	116
B.9	Proof of Theorem 4.4 . . . . .	116
B.10	Proof of Theorem 4.5 . . . . .	117
B.11	Proof of Proposition 4.7 . . . . .	118
B.12	Proof of Proposition 4.8 . . . . .	119
B.13	Proof of Proposition 4.9 . . . . .	120
B.14	Proof of Proposition 4.10 . . . . .	121
B.15	Proof of Lemma 4.6 . . . . .	123
B.16	Proof of Proposition 4.12 . . . . .	123
B.17	Proof of Lemma 4.7 . . . . .	124
B.18	Proof of Theorem 4.6 . . . . .	125

<b>C</b>	<b>Appendix: Whittle’s index policy for minimizing the Age of Information</b>	<b>127</b>
	C.1 Proof of Theorem 5.1 . . . . .	127
	C.2 Proof of Proposition 5.2 . . . . .	129
	C.3 Proof of Theorem 5.3 . . . . .	129
	C.4 Proof of Proposition 5.4 . . . . .	129
	C.5 Proof of Lemma 5.1 . . . . .	136
	C.6 Proof of Lemma 5.2 . . . . .	136
	C.7 Proof of Lemma 5.3 . . . . .	138
	C.8 Proof of Proposition 5.8 . . . . .	138
	C.9 Proof of Proposition 5.9 . . . . .	139
	C.10 Proof of Lemma C.6 . . . . .	141
	C.11 Proof of Proposition 5.10 . . . . .	141
	C.12 Proof of Proposition 5.11 . . . . .	143
	C.13 Proof of Proposition 5.12 . . . . .	145
	C.14 Proof of Proposition 5.13 . . . . .	146
	C.15 Proof of Proposition 5.14 . . . . .	155
	C.16 Proof of Proposition 5.15 . . . . .	158
	C.17 Proof of Proposition 5.16 . . . . .	161
	C.18 Proof of Lemma 5.4 . . . . .	165
	C.19 Proof of Proposition 5.17 . . . . .	167

# List of Figures

1.1	Illustration du temps d'attente et du délai de propagation . . . . .	9
1.2	Modèle de files d'attente . . . . .	10
1.3	Illustration de l'évolution de l'AdI . . . . .	11
1.4	Illustration de l'évolution de l'AdI adaptée au problème d'ordonnancement étudié . . . . .	12
1.5	Modèle du système d'âge . . . . .	13
2.1	Illustration of Queuing and Propagation Delay . . . . .	21
2.2	Queuing System Model . . . . .	21
2.3	Illustration of the AoI evolution . . . . .	23
2.4	Illustration of the AoI evolution adapted to the scheduling problem of interest	24
2.5	Age system model . . . . .	25
4.1	System Model . . . . .	40
4.2	$y^n(W)$ in function of $W$ for different values of $n$ ( $L = 5$ and $R = 10$ ): $W(i)$ indicates the Whittle index at state $i$ . . . . .	50
4.3	Hitting Time of $\Omega_\epsilon(\mathbf{z}^*)$ in function of $N$ : (a) $\mathbf{Z}^N(0) = \mathbf{x}$ , (b) $\mathbf{Z}^N(0) = \mathbf{y}$ .	61
4.4	Performance evaluation of Whittle's Index policy . . . . .	61
4.5	Evaluation of $C_1^{\pi,N}$ and $C_2^{\pi,N}$ in function of $N$ under Policy $\Theta$ and Whittle's Index policy WIP . . . . .	63
4.6	Performance evaluation of Whittle's Index policy . . . . .	63
5.1	The states transitions under a threshold policy $n$ when $L < +\infty$ . . . . .	67
5.2	The states transitions under a threshold policy $n$ when $L = +\infty$ . . . . .	67
5.3	Evolution of $z_i^k(\cdot)$ for different states $i$ in function of $\alpha_1(t)$ and $\alpha_2(t)$ under the Whittle Index Policy (the green and the yellow colors refer to class 1 and 2 respectively) . . . . .	78
5.4	The proportions of users at different states at time $t + T_t$ when $\gamma(t + T_t) = 1$ and $0 < \beta(t + T_t) \leq 1$ . . . . .	80
5.5	The proportions of users at different states at time $t + T_t$ when $\beta(t + T_t) = 1$ and $0 < \gamma(t + T_t) \leq 1$ . . . . .	81
5.6	Graphical representation of $T_{\max}$ . . . . .	82
5.8	Average age per user under Whittle Index Policy . . . . .	89

# List of Tables

5.1	Evaluation of $B_\alpha$ under wide range of channel statistics . . . . .	88
C.1	The expressions of the matrices $Q_k$ for $k \neq m$ and $Q_m$ . . . . .	130
C.2	The expressions of the characteristic polynomials of $Q_k$ . . . . .	131
C.3	The expressions of the characteristic polynomials of $Q_m$ . . . . .	135



# Chapter 1

## Introduction en français

### 1.1 Motivation

Au cours du siècle dernier, le monde a assisté à un changement radical dans la communication grâce à l'apparition de technologies électriques et électromagnétiques telles que les téléphones, les télévisions et les ordinateurs. Cette transformation stupéfiante trouve son origine dans les développements pionniers des communications sans fil de Nikola Tesla, connu pour être l'inventeur de la radio. Les technologies de télécommunication ont dès lors connu une progression exponentielle en termes de performance, d'accessibilité et de quantité, au point qu'aujourd'hui, nous vivons dans un monde entièrement connecté où chaque être humain peut communiquer avec un autre être humain, quel que soit son endroit.

Par ailleurs, selon les données d'intelligence en temps réel de GSMA, le nombre d'utilisateurs d'appareils mobiles dans le monde dépasse aujourd'hui les 5 milliards. Cela représente 66.5% de la population mondiale. D'autre part, grâce à l'émergence de l'IA (intelligence artificielle), les appareils sont devenus plus autonomes et conscients d'eux-mêmes. Cela conduit à l'apparition d'un grand nombre de machines dotées d'identifiants uniques et capables de transférer des données sur un réseau sans nécessiter d'interaction entre humains ou entre ordinateurs. Ce phénomène est mieux connu sous le nom d'IdO (Internet des Objets).

En fait, d'ici la fin de l'année 2030, environ 50 milliards de ces dispositifs IdO seront utilisés dans le monde [1], créant ainsi un réseau massif de dispositifs interconnectés couvrant plusieurs domaines tels que les réseaux de véhicules, les maisons intelligentes, la surveillance de l'environnement, les soins médicaux et de santé, la sécurité et les transports. D'autre part, pour assurer une communication fiable entre deux hôtes finaux distants, nous devons disposer de ressources telles que la bande de fréquences. Cependant, ces ressources radio sont limitées par rapport au nombre massif de dispositifs dans le système de réseau actuel. Dans cette mesure, il faut partager efficacement ces ressources disponibles entre les appareils des utilisateurs et les objets connectés pour répondre aux exigences de service de chaque appareil connecté. Les exigences de chaque dispositif peuvent se manifester différemment : Par exemple, certains appareils implémentent des logiciels qui nécessitent un débit de données élevé (le nombre de paquets transmis par unité de temps). D'autres applications sont plus sensibles aux temps d'attente, comme les jeux vidéo et les vidéoconférences, et exigent alors une faible latence, c'est-à-dire que

le temps entre l'envoi de données par un hôte final et la réception de données par l'autre hôte final doit être aussi faible que possible.

Par ailleurs, d'autres applications de surveillance deviennent populaires et attirent l'attention des chercheurs de nos jours. Dans ces applications, certaines entités centrales ont besoin de recevoir des mises à jour fréquentes d'une information d'intérêt telle que la température, la vitesse et ou la position dans les réseaux de véhicules, pour exécuter une tâche donnée. Néanmoins, pour effectuer l'action appropriée, elle doit acquérir l'information la plus récente et la plus fraîche transmise par les capteurs. Cette notion de fraîcheur peut être quantifiée par l'âge de l'information (AdI). Afin d'offrir les services susmentionnés dans les systèmes sans fil, une branche bien connue des problèmes d'optimisation, à savoir les problèmes d'ordonnancement et d'allocation de ressources, a été développée permettant ainsi de répondre aux exigences de tels services.

Le problème de l'ordonnancement a été largement étudié dans le passé, et plusieurs politiques d'allocation ont été développées pour différents contextes dans différents domaines. Plus précisément, dans le domaine des Télécommunications et en particulier dans le domaine des réseaux sans fil, plusieurs politiques d'ordonnancement sont proposées pour obtenir une bonne performance des réseaux et répondre aux exigences de service. En outre, comme il a été mentionné précédemment, l'émergence de l'IdO et la croissance considérable des appareils des utilisateurs dans les réseaux 5G nécessitent le développement des politiques d'allocation des ressources pour optimiser certaines métriques pertinentes telles que le débit, la latence, le temps d'attente et l'âge de l'information lorsqu'il y a un nombre massif de machines. Dans ce contexte, nous abordons dans notre thèse les problèmes d'ordonnancement des utilisateurs et des canaux et nous nous focalisons sur la minimisation du temps d'attente moyen dans les files d'attente et de l'âge moyen de l'information. Plus précisément, nous étudions le problème suivant : Comment allouer  $K$  ressources à  $N$  utilisateurs de manière à minimiser une certaine métrique, en particulier **Age de l'information** et **Temps d'attente**.

## 1.2 Ordonnancement basé sur le temps d'attente

Aujourd'hui, surtout après les perturbations sociales et économiques généralisées causées par la pandémie mondiale COVID-19, une solution alternative a été mise en place pour éviter l'interruption de la communication entre les personnes, notamment dans les entreprises et les institutions publiques. En effet, des technologies telles que la VoIP, la téléconférence et la télé-présence robotique ont été adoptées pour assurer une communication en temps réel entre les collaborateurs et les membres de la famille. Ces technologies permettent un échange quasi simultané d'informations entre l'émetteur et le récepteur.

Dans cette mesure, comme nous vivons dans un monde où un nombre massif de machines sont connectées, il s'avère pertinent de partager la bande passante du canal entre les utilisateurs de manière optimale afin de réduire la latence. Cette métrique peut être décomposée en deux variables : le temps d'attente dans la file d'attente et le délai de propagation. La première concerne la couche logique, elle correspond au temps entre l'arrivée du paquet dans la file d'attente de l'utilisateur et sa sortie de celle-ci. Quant au délai de propagation, il correspond au temps que met le signal pour arriver au récepteur. Dans cette thèse, nous nous concentrons davantage sur le contexte d'ordonnancement qui minimisent la longueur moyenne de la file d'attente (et donc le temps d'attente moyen

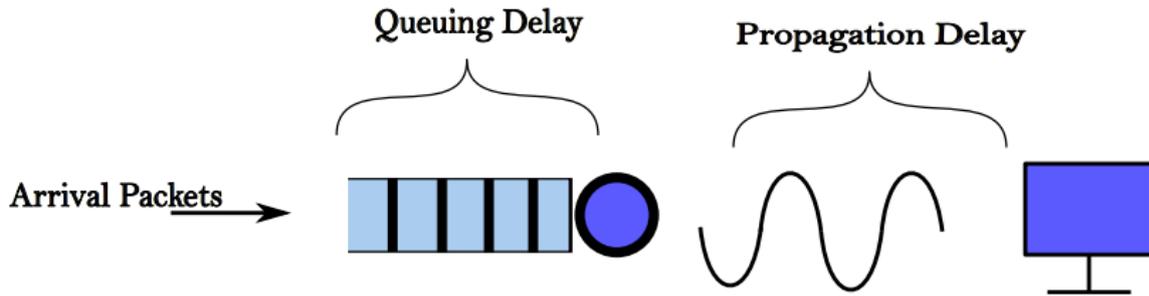


Figure 1.1 – Illustration du temps d’attente et du délai de propagation

de la file) dans un modèle de réseau en temps discret. Ceci dit, le délai de propagation sera normalisé à l’unité de temps. Plus précisément, notre objectif est de trouver une politique d’ordonnancement qui alloue les canaux disponibles à un sous-ensemble donné d’utilisateurs de manière à minimiser la longueur moyenne des files d’attente des utilisateurs, ce qui résulte, selon la loi de Little [2] en la minimisation du temps d’attente moyen. Par conséquent, pour traiter ce problème, nous devons d’abord étudier l’évolution de la file d’attente pour chaque utilisateur. Pour cela, nous donnons dans ce qui suit un bref aperçu du modèle de file d’attente discrète considéré dans notre thèse en relation avec le problème qui nous intéresse :

Notre système de files d’attente est composé de  $N$  files d’attente et de  $K$  canaux. Le nombre de canaux est inférieur au nombre de files d’attente. De plus, chaque file d’attente a un processus d’entrée et de sortie. Le processus d’entrée indique l’arrivée des paquets dans la file, tandis que le processus de sortie désigne le départ des paquets de la file. Dans le cas discret, le processus d’arrivée suit un processus discret iid (indépendant et identiquement distribué). Plus précisément, le nombre de paquets arrivés à l’instant actuel ne dépend pas de celui de l’instant précédent. De même, le processus de départ est un processus iid discret dans le temps. Dans cette mesure, en désignant par  $Q(t)$ ,  $A(t)$ ,  $D(t)$ , la longueur de la file d’attente, le processus d’arrivée, et le processus de départ respectivement au temps  $t$ , la longueur de la file d’attente évolue de façon stochastique comme suit :

$$Q(t + 1) = \max\{Q(t) - D(t), 0\} + A(t)$$

En se basant sur l’équation ci-dessus, le processus  $Q(t)$  peut être modélisé comme une chaîne de Markov puisque la valeur de  $Q(\cdot)$  à l’instant suivant ne dépend que de l’état actuel. Cette propriété est mieux connue comme loi sans mémoire. Dans cette mesure, notre objectif est de sélectionner à chaque instant,  $K$  parmi  $N$  utilisateurs auxquels les  $K$  canaux sont alloués de manière à minimiser l’expectation de la moyenne totale des files d’attente de tous les utilisateurs (Figure 1.2). L’expression explicite du problème susmentionné est fournie dans le chapitre 4.

### 1.3 Ordonnancement basé sur l’âge

Il est largement reconnu que l’IdO apportera une amélioration impressionnante dans divers domaines et secteurs de nos activités quotidiennes. Dans cette mesure, il est en-

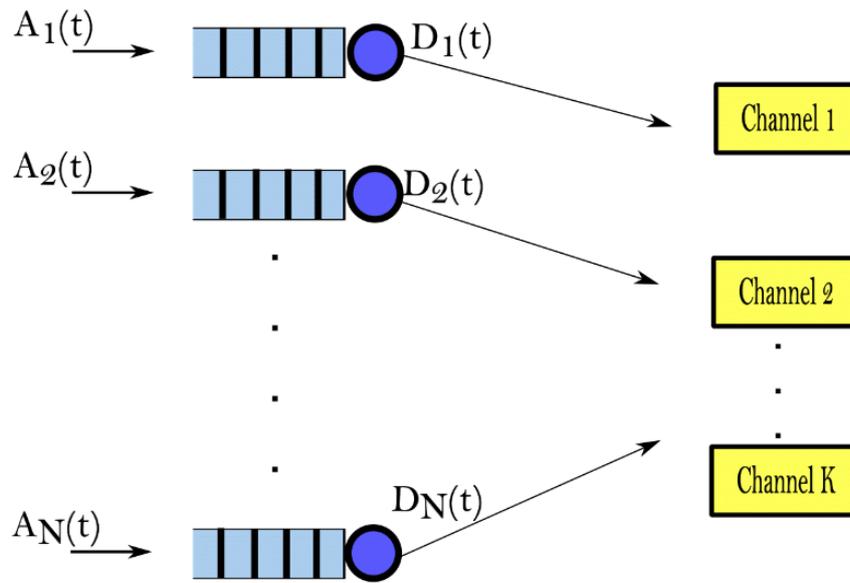


Figure 1.2 – Modèle de files d’attente

visagé qu’un monde numérique futur soit composé de nombreux capteurs et dispositifs connectés sans fil et à faible coût. Cela permettra aux moniteurs d’exécuter certaines opérations en exploitant les informations acquises par des capteurs distants. Cependant, l’information du côté des moniteurs doit être aussi fraîche que possible pour effectuer les opérations appropriées. Cette notion de fraîcheur peut être capturée, comme il a été mentionné précédemment, par la notion d’âge de l’information. En effet, le concept d’âge de l’information (AdI) a été introduit pour la première fois en 2011 dans le but de quantifier la fraîcheur des informations dont nous disposons sur l’état d’une donnée d’intérêt. Plus précisément, l’AdI peut être considérée comme la durée qui sépare l’instant de génération du dernier paquet reçu avec succès et l’heure actuelle. L’AdI a attiré de nombreux chercheurs (par exemple, [3–6]) avec plus de 50 publications à ce jour en raison de sa bonne adéquation avec le concept de fraîcheur de l’information. En effet, on peut remarquer que les métriques de débit, temps d’attente et la latence ne capturent pas la notion d’intemporalité ou de fraîcheur de l’information comme le fait l’AdI. Pour souligner ce fait, nous expliquons les deux approches suivantes où la source transmet des paquets sous forme des mises à jour d’une certaine information.

- Maximisation du débit:

Pour maximiser le débit d’un hôte final donné, ou explicitement pour maximiser la quantité de données envoyées par la source dans une unité de temps, nous devons augmenter autant que possible le taux de génération de paquets pour tirer le meilleur parti de la capacité du canal. L’inconvénient de cette méthode est qu’un délai de retard sera encouru dans la file du canal. Par conséquent, le moniteur recevra des paquets pour lesquels le temps de livraison est considérablement plus grand que leur temps de génération. Donc, cette approche ne permet pas de satisfaire la propriété de ponctualité du système de mise à jour

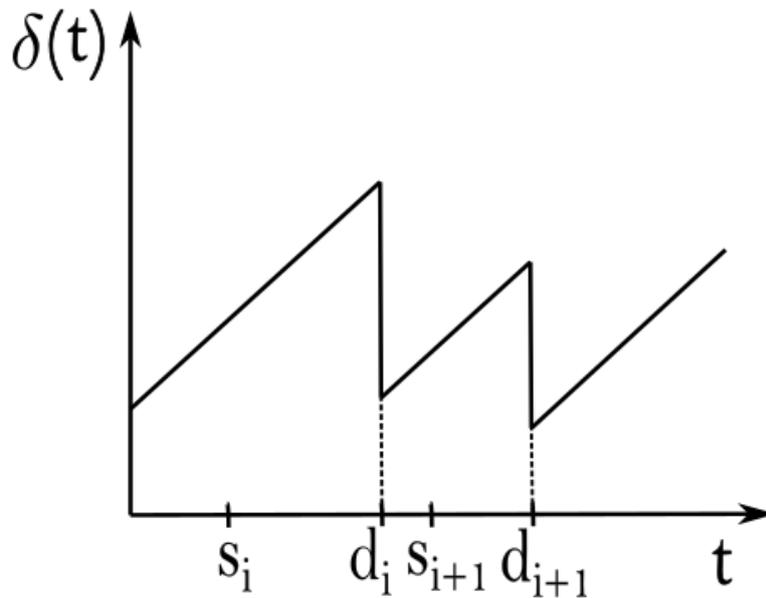


Figure 1.3 – Illustration de l'évolution de l'AdI

- Minimisation du délai de transmission:

Nous expliquons maintenant pourquoi la minimisation du temps d'attente ne peut pas capturer la notion de fraîcheur de l'information. En fait, la minimisation du temps d'attente des paquets peut être obtenue en diminuant le taux de génération des paquets pour alléger la charge du système. De cette façon, le temps entre la génération de la mise à jour de l'état et sa réception par le moniteur sera réduit. Cependant, le moniteur recevra des informations obsolètes sur l'état du système puisque le taux de génération des mises à jour est faible.

C'est pour cette raison que la métrique de l'âge de l'information a été créée pour optimiser le système de mise à jour des statuts. Mathématiquement, l'AdI est formulé comme suit:

$$\delta(t) = t - \max_i \{s_i : d_i \leq t\}$$

où  $s_i$  et  $d_i$  désignent respectivement le temps de génération du  $i$ -ième paquet de mise à jour de l'information en question et son temps de livraison au moniteur. D'après cette définition, on peut remarquer qu'effectivement  $\delta(t)$  capture bien la notion de fraîcheur puisqu'une petite valeur de  $\delta(t)$  signifie que le moniteur possède une mise à jour récente, et si  $\delta(t)$  est grand, le moniteur possède une mise à jour obsolète.

Ces mises à jour peuvent être générées soit à la volonté de la source-captteur, soit spontanément en fonction des facteurs externes tels que les changements climatiques ou les variations de vitesse.

Dans cette mesure, en considérant le cadre du problème d'ordonnancement, dans le scénario où la source distante peut générer des mises à jour à n'importe quel moment, il semble que pour éviter une perte inutile d'énergie, la source ou le capteur ne devrait pas générer de mise à jour s'il n'est pas autorisé à transmettre. Par conséquent, la manière

pertinente et optimale de minimiser autant que possible l'énergie consommée par les capteurs est de produire des paquets uniquement à la demande du moniteur, c'est-à-dire lorsque celui-ci alloue le canal au capteur en question. Cela conduit également à réduire l'AdI moyen, puisque le délai entre la génération et la réception du paquet transportant l'information sur le processus d'intérêt est limité uniquement au délai encouru dans la file du canal. De plus, en supposant qu'il n'y a pas de délai supplémentaire causé par le canal, à l'exception de celui de la propagation du signal qui est normalisé à un, alors  $s_i$  et  $d_i$  qui désignent le temps de génération et le temps de livraison du  $i$ -ième paquet de mise à jour reçu avec succès par le moniteur respectivement, vérifient :

$$d_i = s_i + 1$$

Si le paquet est perdu pendant la transmission, la mise à jour correspondante est rejetée. Par conséquent, il faut donner une nouvelle formulation mathématique qui correspond au problème d'ordonnancement étudié dans notre thèse au chapitre 5. qui sera décrit en détail dans le chapitre 5.1.2. L'évolution de l'AdI est :

$$\delta(t) = t - \max_i \{s_i : s_i \leq t - 1\} \quad (1.1)$$

À la lumière de ce fait, on peut observer sur la figure 1.4 que lorsque une nouvelle mises à jour de l'information en question est générée, ou de manière équivalente, lorsque le canal est alloué au capteur concerné, la valeur de l'âge passe à 1 à l'instant suivant en supposant que la transmission est réussie.

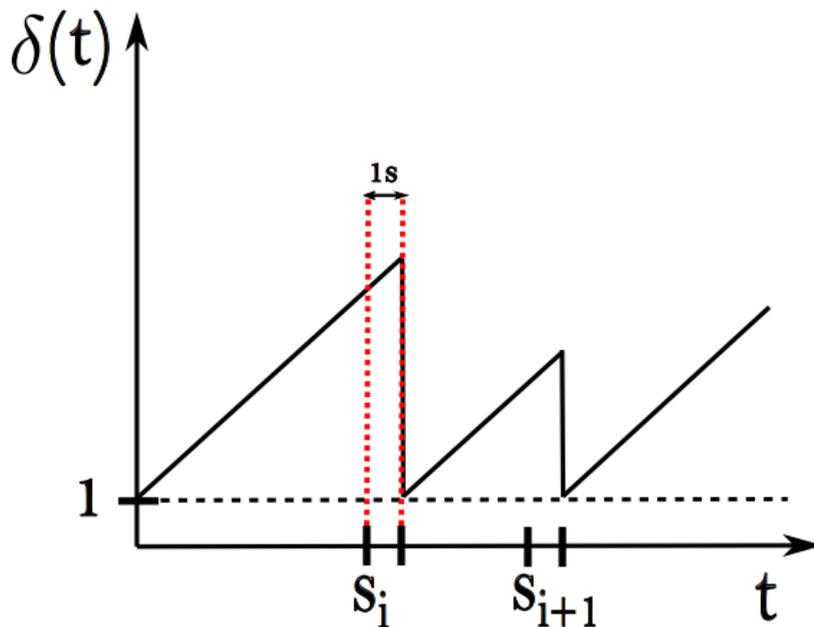


Figure 1.4 – Illustration de l'évolution de l'AdI adaptée au problème d'ordonnancement étudié

Comme le nombre d'hôtes finaux connectés dans les réseaux a considérablement augmenté alors que les ressources disponibles telles que les canaux de transmission sont toujours limitées, il est intéressant d'étudier les problèmes d'allocation des ressources pour minimiser

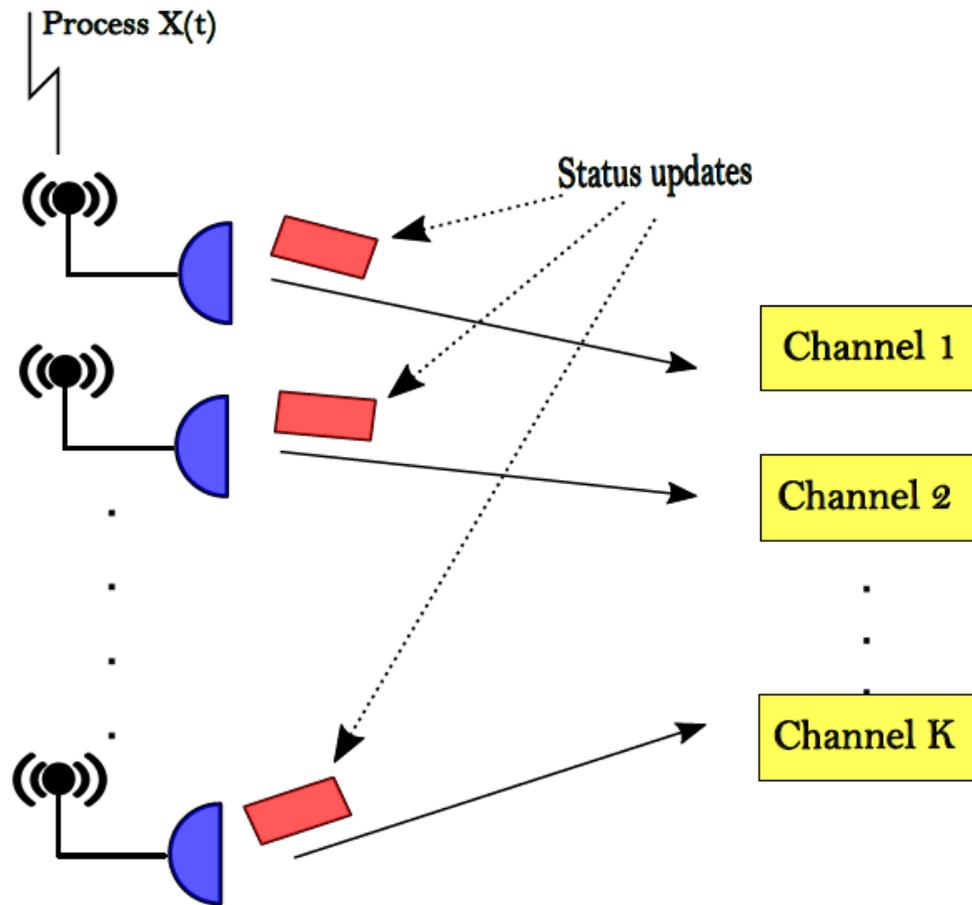


Figure 1.5 – Modèle du système d'âge

la métrique AdI dans un grand système de réseau. Dans cette optique, nous étudions dans notre thèse le problème suivant : comment allouer les  $K$  canaux aux  $N$  utilisateurs de manière à minimiser l'âge moyen total de l'information de tous les utilisateurs du système (Figure 1.5). La dérivation mathématique du problème susmentionné est donnée au chapitre 5.

## 1.4 Etat de l'Art

### 1.4.1 Problème d'ordonnancement: Minimisation du temps d'attente

Plusieurs travaux ont étudié les problèmes d'ordonnancement pour minimiser la métrique de temps d'attente (par exemple : [7–12]). Dans la suite, nous citons quelques travaux pertinents dans ce domaine particulier.

#### Ordonnancement tenant compte de l'énergie

L'économie d'énergie fait l'objet d'une grande attention dans les réseaux 5G. Cependant, nous savons qu'il existe un compromis entre l'économie d'énergie et les garanties de per-

formance, notamment les délais de bout en bout.

Dans cette optique, [13] étudie la transmission à puissance minimale sur des canaux sans fil avec des contraintes sur le temps d'attente moyen. Ils proposent un ordonnanceur à faible complexité qui a une performance quasi-optimale.

Concernant l'énergie qui permet à une machine de rester active, qu'elle soit en mode transmission ou en mode veille, on peut considérer que la meilleure façon d'économiser cette énergie consommée est d'éteindre les machines. Cependant, la transition entre les modes actif et passif consomme une énergie considérable et augmente le délai. [14] étudie ce problème dans un réseau multi-sauts et propose pour une topologie de ligne ainsi que pour une topologie arbitraire, une politique d'ordonnement qui est quasi-optimale en ce qui concerne la métrique du délai pour une période active minimale par élément du réseau.

### Algorithmes d'apprentissage stochastiques

Un nombre important de travaux dans la littérature développent des stratégies d'allocation à l'aide d'algorithmes d'apprentissage stochastiques [15–21] afin d'optimiser le temps d'attente dans les files des utilisateurs dans le contexte de problèmes d'ordonnement. Cependant, les cadres des processus de décision de Markov (PDM) souffrent du fléau de la dimensionnalité, ce qui conduit à des stratégies d'allocation des ressources complexes. Par exemple, [12] étudie l'ordonnement des ressources pour la minimisation du temps d'attente dans les systèmes informatiques cellulaires multi-serveurs en périphérie de réseau. Les auteurs proposent une solution basée sur une nouvelle fonction de Lyapunov appelée fonction de Lyapunov basée sur le temps d'attente. Plus précisément, ils conçoivent un algorithme d'ordonnement basé sur cette nouvelle fonction afin de dériver une solution approximative qui minimise à la fois la file d'attente et le délai de propagation. Ils établissent ensuite que l'algorithme proposé donne de bonnes performances par rapport à l'algorithme d'ordonnement traditionnel basé sur la méthode de Lyapunov. Cependant, l'algorithme dérivé reste sous optimal et sa complexité est considérablement élevée.

Dans [18, 22], les auteurs tentent de minimiser le temps d'attente moyen dans les files des utilisateurs en utilisant le processus de décision de Markov (PDM) et des outils d'apprentissage stochastique. Cette méthode a également été adopté dans [15] pour traiter le problème de l'allocation de puissance dans un système OFDM (Orthogonal Frequency Division Multiplexing) avec pour objectif de minimiser le temps d'attente moyen des paquets des utilisateurs dans les files d'attente. Cependant, la solution développée nécessite une mémoire et une complexité de calcul élevées.

### Politique de Whittle index pour minimiser le temps d'attente

Les politiques basées sur l'indice de Whittle ont été utilisées/développées dans les réseaux sans fil dans le cadre de problèmes d'ordonnement en raison de leur faible complexité et de leurs bonnes performances dans la plupart des cas : [23–36]. Dans [24], une dérivation des valeurs de l'indice de Whittle pour un modèle M/M/1 multi-classe simple a été considérée (où un seul utilisateur peut être servi). Cependant, l'optimalité de la politique d'indice de Whittle obtenue n'a pas été prouvée. Les auteurs dans [37] ont considéré le problème de l'ordonnement des projets/travaux dans lequel un effort est alloué à un

nombre fixe de projets. La performance d'une politique basée sur l'indice de Whittle a été analysée dans un modèle à temps continu.

Dans [38], une politique basée sur l'indice de Whittle a été dérivée pour un système de file d'attente composé de plusieurs classes et serveurs où l'arrivée et le départ des paquets évoluent selon un processus de naissance et de mort. Dans [39], une politique d'indexation optimale appelée loi  $c\mu$  généralisée ( $Gc\mu$ ) est développée dans le contexte d'un régime de trafic intense avec un coût de temps d'attente convexe. En outre, contrairement à la politique de la règle  $c\mu$ , [40] établit l'optimalité de la règle  $c\mu$ -généralisée ( $Gc\mu$ ) même avec plusieurs serveurs.

Dans [24], les auteurs calculent la politique d'indexation de Whittle pour un système de file d'attente comportant plusieurs classes avec des fonctions de coût générales. Nous notons que pour ces travaux cités, contrairement à notre modèle, le temps a été considéré comme étant continu.

### 1.4.2 Problème d'ordonnancement: Minimisation de l'Age

Parmi les premiers travaux qui étudient le problème de minimisation de l'AdI, on trouve celui de Kaul, Gruteser et Yates [4]. Ils considèrent un modèle de système où une source transmet des paquets contenant des mises à jour d'une information particulière à une destination. Le canal est représenté par une simple file d'attente. Dans ce travail, les auteurs montrent que ni l'envoi de mises à jour à une forte cadence ni l'envoi de mises à jour à une faible cadence ne permettent d'avoir une livraison opportune des paquets transmis. En effet, la solution optimale n'est pas triviale, comme on peut le prévoir.

Dans [41], les auteurs étudient l'impact de la salle d'attente sur l'AdI dans un système de file d'attente où plusieurs flux d'information partagent un service commun avec une seule salle d'attente qui peut accueillir au plus un paquet. Ces flux ont des priorités différentes. Les auteurs ont trouvé le taux d'arrivée optimal des mises à jour qui minimise l'âge moyen de chaque flux.

Dans le cadre des problèmes d'ordonnancement, de nombreux travaux (par exemple, [42–48]) ont étudié ce type de problèmes afin de minimiser l'âge de l'information dans des contextes variés. Dans cette optique, nous citons dans ce qui suit quelques travaux intéressants liés à notre domaine d'études.

#### Ordonnancement tenant compte de l'énergie

[47] étudie un scénario de détection en temps réel sous contraintes d'énergie et de batterie. Les auteurs trouvent une politique de seuil optimale en terme d'AdI en fonction de l'énergie et des états d'âge actuels estimés. Un autre travail intéressant dans ce domaine est [48], où les auteurs étudient un problème d'ordonnancement pour minimiser l'AdI en considérant que le taux de mise à jour de la source ne peut pas dépasser une certaine limite prédéfinie en raison de limitations énergétiques. Ils fournissent une dérivation analytique de la solution optimale lorsque les statistiques du canal sont connues.

#### Algorithmes d'apprentissage stochastiques

De nombreux algorithmes ont été développés pour trouver l'ordonnancement optimal par rapport à l'AdI en se basant sur des outils d'apprentissage stochastiques. Par exemple, [49]

propose un algorithme d'apprentissage par renforcement à coût moyen qui apprend les paramètres du système, à savoir les statistiques de canal et de récolte d'énergie, afin d'obtenir la politique optimale.

Dans [50], les auteurs considèrent le problème de l'ordonnancement de plusieurs flux desservis par un seul serveur. Ils utilisent des méthodes de gradients et de Q-Learning pour obtenir des décisions d'ordonnancement qui résistent aux conditions du réseau et aux processus d'arrivée des paquets.

D'autre part, une approche d'approximation du processus de décision de Markov (PDM) a été adoptée pour proposer des algorithmes d'ordonnancement en ligne et hors ligne dans le cas asymétrique dans [44], qui s'avèrent asymptotiquement optimaux.

### Politique de Whittle index pour minimiser AdI

Comme indiqué dans le chapitre 1.4.1, la politique de l'indice de Whittle a été largement appliquée au traitement des problèmes d'ordonnancement en raison de ses avantages par rapport aux autres politiques d'ordonnancement, à savoir sa faible complexité et ses bonnes performances. Plus précisément, la politique de l'indice de Whittle est utilisée dans le cadre du problème du bandit sans repos qui sera détaillé au chapitre 3. L'une des instances les plus connues de ce problème est : Comment allouer  $M$  ressources à  $N$  utilisateurs ( $M$  inférieur à  $N$ ) de manière à minimiser une métrique donnée qui évolue de manière stochastique. [43] traite ce problème susmentionné. Sachant que le RBP (Restless Bandit Problem) est un cas particulier de PDM (Problème de décision de Markov) dont la solution est connue pour être inatteignable dans de nombreux cas, [43] ont prouvé qu'un algorithme glouton est optimal lorsque les utilisateurs ont des statistiques de canal identiques, tandis que pour le cas asymétrique, ils proposent comme solution la politique d'indexation de Whittle qui reste sous-optimale. Dans [45], les auteurs cherchent à minimiser l'AdI moyen total, sous des arrivées aléatoires en adoptant une politique basée sur l'indice de Whittle. En conséquence, ils dérivent une forme explicite et simple de la politique de l'indice de Whittle. Ils donnent en outre quelques résultats numériques pour montrer la performance de la politique proposée. Cependant, ils ne prouvent pas son optimalité. Dans cette mesure, nous adoptons dans notre travail, cette politique performante, et nous établissons en outre son optimalité dans le régime asymptotique lorsque le nombre d'utilisateurs augmente.

## 1.5 Contributions et Aperçu de la thèse

Cette thèse est structurée en trois chapitres. Plus précisément :

- Chapitre 3: Dans ce chapitre, nous expliquons en détail la méthodologie appliquée pour obtenir la politique de l'indice de Whittle (WIP). Plus précisément, nous expliquons en détail les étapes pour obtenir les indices de Whittle et nous exposons la méthode utilisée pour montrer l'optimalité de WIP dans le régime multi-utilisateurs.
- Chapitre 4: Contrairement à [24,37–40], nous abordons dans ce chapitre un problème d'ordonnancement en considérant **un modèle en temps discret**. Plus précisément, nous étudions le problème de minimisation de temps d'attente dans un modèle de système de file d'attente en temps discret. Dans cette mesure, nous étudions

un problème d'allocation de ressources dans lequel une station de base alloue  $M$  canaux à  $N$  files d'attente d'utilisateurs à chaque instant en fonction de la politique d'ordonnancement considérée. Les files d'attente évoluent selon un processus de Markov. En outre, les utilisateurs ou les files d'attente <sup>1</sup> sont divisés en classes en fonction des paramètres du système, à savoir le taux d'arrivée des paquets et le taux de transmission des paquets, qui diffèrent d'une classe à l'autre. Notre objectif est de trouver la politique adéquate qui minimise la longueur moyenne totale attendue de la file d'attente de tous les utilisateurs sur le long terme. Le problème susmentionné peut être considéré comme un problème de bandit sans repos dont la solution est connue pour être inaccessible. Pour cela, nous appliquons la politique peu complexe et très performante, la politique de l'indice de Whittle (WIP) à notre modèle spécifique. Nous pouvons résumer nos contributions en deux points :

- Nous dérivons la politique de l'indice de Whittle pour deux scénarios : files d'attente de taille infinie, et files d'attente de taille restreinte (inférieure au taux de transmission).
  - Nous établissons l'optimalité locale et globale du WIP dans le régime asymptotique lorsque le nombre d'utilisateurs est très grand.
- Chapitre 5: Dans ce chapitre, nous appliquons la politique de l'indice de Whittle dans le contexte de l'âge de l'information (AdI). Nous considérons un problème d'ordonnancement avec des canaux non fiables. Nous considérons plusieurs classes d'utilisateurs telles que chaque classe est caractérisée par ses propres statistiques de canal. Par conséquent, notre objectif est de minimiser l'âge moyen total attendu à long terme de tous les utilisateurs. Notre principale contribution dans ce travail est la preuve de l'optimalité asymptotique pour les deux scénarios :
    - AdI évolue dans un espace d'état fini.
    - AdI évolue dans un espace d'état infini.

Pour le premier scénario, nous fournissons une analyse rigoureuse afin d'établir l'optimalité locale asymptotique basée sur certaines techniques mathématiques. Pour le second scénario, nous adoptons une approche innovante et originale pour établir l'optimalité globale asymptotique de la politique de l'indice de Whittle. Cette approche repose essentiellement sur le critère de Cauchy : au lieu de prouver qu'une fonction donnée  $f(\cdot)$  converge vers un point fixe, nous montrons que les termes de  $f(\cdot)$  se rapprochent lorsque  $t$  croît. Cette méthode sera détaillée dans le chapitre 5. Enfin, nous fournissons quelques résultats numériques qui confirment notre analyse technique.

---

<sup>1</sup>Les termes utilisateurs et files d'attente sont interchangeable dans ce chapitre



# Chapter 2

## Introduction

### 2.1 Motivation

In the last century, the world has witnessed a radical change in communication thanks to the apparition of electrical and electromagnetic technologies such as phones, televisions, and computers. This astounding transformation has its origin in the pioneering developments in wireless communications by Nikola Tesla, who is known to be the inventor of the radio. The telecommunication technologies have thenceforth known an exponential advancement in terms of performance, reachability, and quantity to the extent that today, we live in a fully connected world where each human-being can communicate with the other human being regardless of their locations. Besides, according to GSMA real-time intelligence data, the number of mobile device users worldwide today surpasses 5 billion. That is 66.5% of the world's population.

On the other side, due to the emergence of AI (Artificial Intelligence), the devices became more autonomous and self-aware. That leads to the apparition of a huge number of machines provided with unique identifiers and can transfer data over a network without requiring human-to-human or human-to-computer interaction. This is better known by IoT (Internet of Things). In fact, by the end of 2030, around 50 billion of these IoT devices will be in use around the world [1], creating a massive web of interconnected devices covering several domains such as vehicular network, smart homes, environmental Monitoring, Medical and health-care, security, transportation.

On the other hand, to ensure reliable communication between two distant end-hosts, we need to afford resources such as frequency band. However, these radio resources are limited compared to the massive number of devices in the network system nowadays. To that extent, one should effectively share these available resources between users' devices and connected objects to meet the service requirement for each connected device. Each device's requirements can be manifested differently: For instance, some devices implement software that needs a high data rate (the number of transmitted packets per time unit). Other applications are more delay sensitives, such as video games and video conferencing, and then require low latency, i.e., the time between sending data by one end-host and receiving data by the other end-host should be as low as possible. Besides, other monitoring applications are becoming popular and attract the attention of researchers nowadays. In such applications, some central entities need to receive frequent updates about an information of interest such as temperature, velocity and position in vehicular networks,

to execute a given task. Nevertheless, to perform the appropriate action, it needs to acquire the up-to-date and the freshest information from the sensors. This notion of freshness can be captured by Age of Information (AoI).

In order to offer the aforementioned services in wireless systems, a well known branch of optimization problems, namely scheduling and resource allocation problems, has been developed allowing thus to meet the requirements of such services.

The scheduling problem has been widely studied in the past, and several allocation policies have been developed for various contexts in different fields. Specifically, in the field of Telecommunication and precisely in wireless network domain, several scheduling policies are proposed to achieve a good performance of networks and meet the service requirements. Moreover, as it was mentioned before, the emergence of IoT and the considerable growth of users' handsets in 5G networks requires the development of resource allocation frameworks to optimize some relevant metrics such as throughput, latency, queuing delay, and age of information when there are a massive number of machines. In this context, we tackle in our thesis users and channels scheduling problems and we focus on minimizing the average delay and the average age of information. More precisely, we study the following problem: How to allocate  $K$  resources to  $N$  users in such a way to minimize a certain metrics, in particular **Age Of Information** and **Queuing Delay**.

## 2.2 Delay-based Scheduling

Today, especially after the widespread social and economic disruption caused by the global pandemic COVID-19, an alternative solution has taken place to avoid interruption of communication between people, particularly in companies and state institutions. Indeed, technologies such as VoIP, teleconferencing, and robotic telepresence have been adopted to ensure real-time communication between the collaborators and family members. These technologies allow a near simultaneous exchange of information from the sender to the receiver in a connection with negligible latency. To that extent, as we live in a world where a massive number of machines are connected, it turns out to be relevant to share the channel bandwidth between users in an optimal way in order to reduce the latency. This metric can be decomposed into two variables: queuing delay and propagation delay. The first one concerns the logical layer, it refers to the time between the packet arrival in user's queue and its departure from it. As for the propagation delay, it refers to the amount of time it takes for the signal to travel from the sender to the receiver. In this thesis, we focus more on scheduling frameworks that minimize the average queuing length (and hence the average queuing delay) in a discrete time network model. Having said that, the propagation delay will be normalized to the time unit. Specifically, our objective is to find a scheduling policy that allocates the available channels to a given subset of users in such a way to minimize the average queue length of the users which results, according to Little's Law [2] in the minimization of the average delay. Therefore, to deal with this problem, we need first to investigate the evolution of the queue for each user. To proceed so, we give in the following a brief insight about the discrete queuing model considered in our thesis related to our problem of interest:

Our queueing system is composed by  $N$  queues and  $K$  channels. The number of channels is less than the number of queues. Furthermore, each queue has an input and output process. The input process indicates the packets' arrival to the queue, while the output process

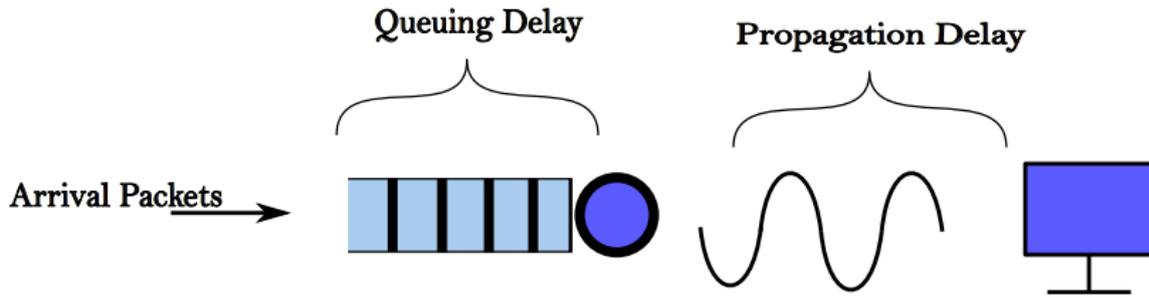


Figure 2.1 – Illustration of Queuing and Propagation Delay

designates the packets' departure from the queue. In the discrete case, the arrival process follows a discrete iid (independent and identically distributed) process. Specifically, the number of arrived packets at the current time doesn't depend on that at the previous time slot. Likewise, the departure process is a discrete iid process over time. To that extent, denoting by  $Q(t)$ ,  $A(t)$ ,  $D(t)$ , the queue length, the arrival process, and the departure process respectively at time  $t$ , the queue or the buffer length evolves stochastically as follows:

$$Q(t + 1) = \max\{Q(t) - D(t), 0\} + A(t)$$

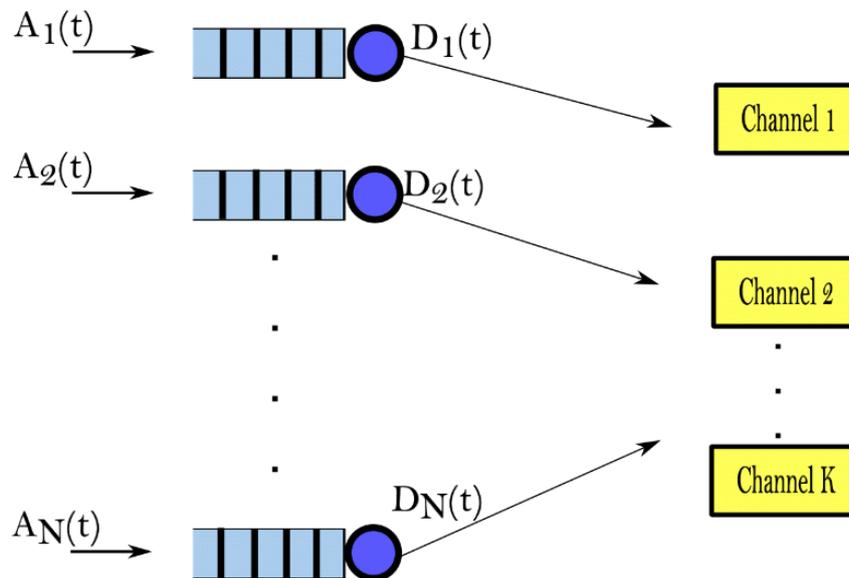


Figure 2.2 – Queuing System Model

Based on the above equation, the process  $Q(t)$  can be modeled as a Markov chain since the value of  $Q(\cdot)$  at the next time slot depends only on the current state. This property is better known as memoryless. To that extent, our goal is to select per each time slot  $K$  among  $N$  users to which the  $K$  channels are allocated in such a way to minimize the

total expected average queues of all users. The explicit expression of the aforementioned problem is provided in Chapter 4.

## 2.3 Age-based Scheduling

It is widely recognized that IoT will bring an impressive enhancement in various areas and domains of our daily activities. To that extent, it is envisioned that a future digital world will be composed of numerous wireless low-cost connected sensors and devices. This will allow to the monitors to execute some operations exploiting the information acquired remotely from the sensors. Meanwhile, the information at the monitors side should be as fresh as possible to perform the appropriate and convenient operations. Accordingly, as it was mentioned previously, the notion of freshness can be captured by the notion of AoI. Indeed, the concept of Age of Information (AoI) was introduced the first time in 2011 in [51] to quantify the freshness of the information that we have about the status of a given data of interest. More specifically, AoI can be viewed as the duration which separates the generation of the last successfully received packet's time-stamp and the current time.

AoI has attracted many researchers (for instance, [3–6]) due to its well-fitting with the concept of freshness of information. Indeed, one can remark that the metrics of throughput, delay, and latency don't capture the notion of timelessness or freshness of the information as AoI does. To highlight this fact, we explain the two following approaches where the source transmits packets as status updates.

- **Throughput maximization:**  
To maximize the throughput of a given end-host, or explicitly to maximize the amount of data sent by the source within a unit of time, we need to increase as much as possible the generation rate of packets to make the most of the canal capacity. The shortcomings of this is high delays will be incurred in channel buffer. Therefore, the monitor will receive packets for which the delivery time is considerably larger than their generation time. As a consequence, this approach falls short in satisfying the timeliness property of the status updates system.
- **Delay minimization:**  
We now explain why the delay minimization cannot capture the notion of freshness of information. In fact, minimizing the packets' delay can be achieved by decreasing the generation rate of packets to alleviate the burden on the system. By doing so, the time between the generation of the status update and its reception by the monitor will be reduced. However, the monitor will receive outdated status information about the system since the updates generation rate is low.

For this reason, the Age of Information metric was arisen to optimize the status updates system. Mathematically, AoI is formulated as follows:

$$\delta(t) = t - \max_i \{s_i : d_i \leq t\}$$

where  $s_i$  and  $d_i$  refer to the generation time of the  $i$ -th status update packet and the delivery time to the monitor respectively. From this definition, one can notice that effectively  $\delta(t)$  captures well the notion of the freshness since a small value of  $\delta(t)$  implies

that the monitor possess a fresh status update, and if  $\delta(t)$  is large, the monitor possess an outdated status update.

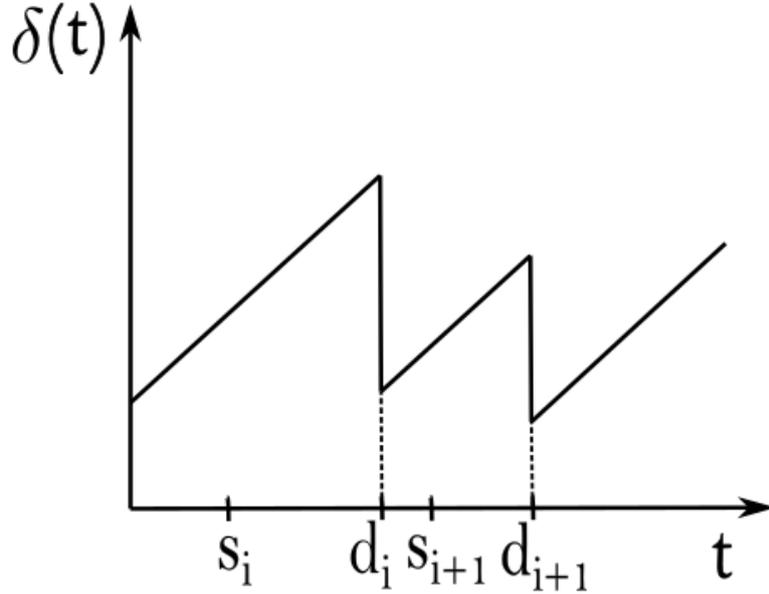


Figure 2.3 – Illustration of the AoI evolution

These status updates can be either generated at the will of the source-sensor or spontaneously depending on external factors such as climate change or velocity variation.

To that extent, considering the scheduling problem's framework, in the scenario where the remote source can generate status updates at any desired time instance, it sounds that to avoid a useless waste of energy, the source or the sensor should not generate a status update if it is not allowed to transmit. Subsequently, the pertinent and the optimum way to minimize as possible the energy consumed by sensors is to produce packets only at the request of the monitor, i.e., when this later allocates the channel to the sensor in question. This leads also to reduce the average AoI, since the delay between the generation and the reception of the packet carrying the information about the process of interest is restricted only to the delay incurred in the channel buffer. Moreover, assuming that there is no additional delay caused by the channel except that of signal's propagation which is normalized to one, then  $s_i$  and  $d_i$  that refer to the generation time slot and delivery time slot of the  $i$ -th status update packet successfully received by the monitor respectively, verify:

$$d_i = s_i + 1$$

If the packet is lost during the transmission, then the corresponding status update is discarded.

Therefore, one should give a new mathematical formulation that fits with the scheduling problem studied in our thesis in Chapter 5. Accordingly, the evolution of AoI is:

$$\delta(t) = t - \max_i \{s_i : s_i \leq t - 1\} \quad (2.1)$$

In light of that fact, one can observe in figure 2.4 that when the new status updates is generated, or equivalently, when the channel is allocated to the concerned sensor, the value of the age goes to 1 at the next time slot assuming that the transmission is successful.

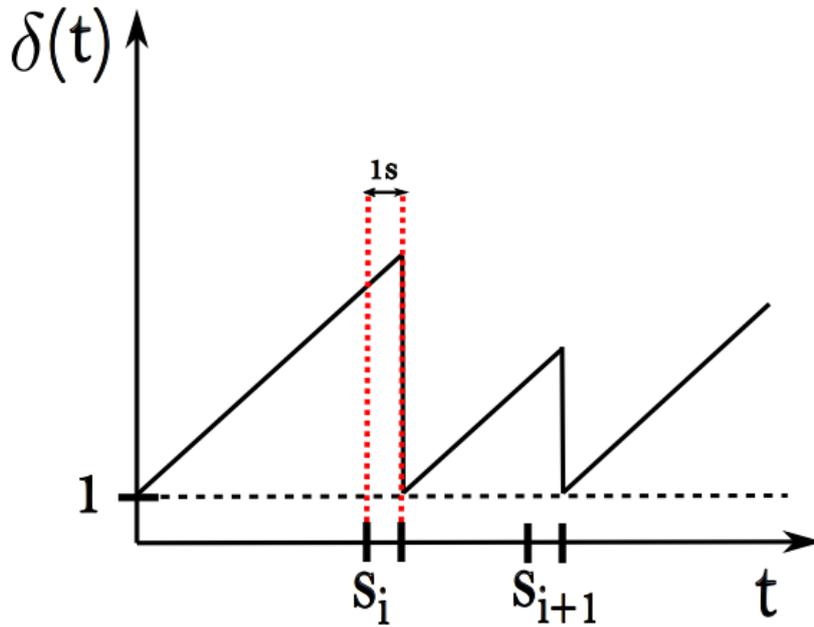


Figure 2.4 – Illustration of the AoI evolution adapted to the scheduling problem of interest

As the number of connected end-hosts in the networks has significantly increased while the available resources such as transmission channels are still limited, it is worthwhile to study the resource allocation problems to minimize AoI metric in a large network system. To that extent, we study in our thesis the following problem: how to allocate the  $K$  channels to the  $N$  users in such a way to minimize the total average age of information of all users of the system. The mathematical derivation of the aforementioned problem is given in Chapter 5.

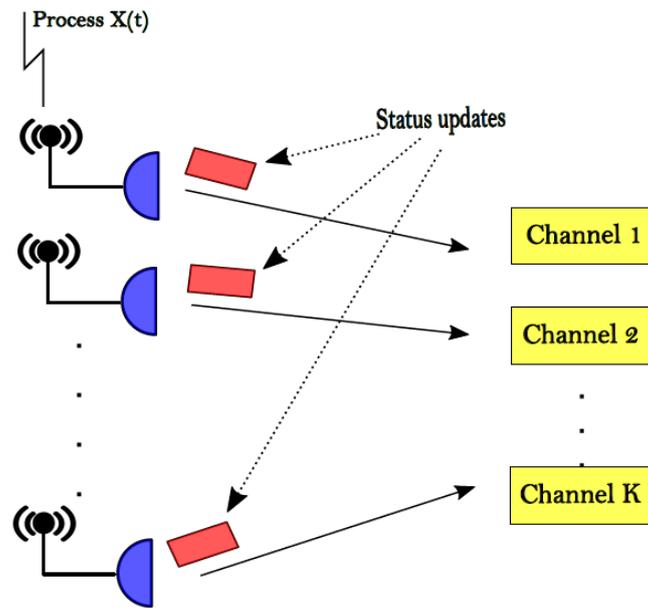


Figure 2.5 – Age system model

## 2.4 State of the Art

### 2.4.1 Scheduling Problem: Delay Minimization

Several works have studied the scheduling problems for minimizing the delay metric (for instance [7–12]). In the sequel, we cite some relevant works in this particular area.

#### Energy-aware scheduling

Energy conservation is getting great attention in 5G networks; however, as we know that there is a trade-off between saving energy and provisioning performance guarantees, especially end-to-end delays.

To that extent, [13] studies minimal power transmission of bursty sources over wireless channels with constraints on mean queuing delay. They propose a low-complexity scheduler that has a near-optimal performance.

Regarding the energy that allows a machine to stay active whether it is in a transmit mode or idle mode, one can consider that the best way to save this power consumed is to turn off the machines. However, transitioning between the active and sleeping modes consumes considerable power and increases the delay. [14] investigates this problem in a multi-hop network and proposes for a line topology as well as for an arbitrary topology, a scheduling policy that is near-optimal with regards to delay metric for a minimum active period per element of the network.

#### Stochastic learning algorithms

Significant number of works in the literature use Markov Decision Process (MDP) frameworks and develop allocation strategies using stochastic learning algorithms [15–21], (e.g.,

by using value iteration, policy iteration, etc.) in order to optimize the delay in users' queues in scheduling problem backgrounds. However, MDP frameworks suffer from the curse of dimensionality, which leads to complex resource allocation strategies. For instance, [12] studies resource scheduling for delay minimization in multi-server cellular edge computing systems. The authors propose a solution based on a new Lyapunov function called the delay-based Lyapunov function. Specifically, they design a scheduling algorithm based on this new function to derive an approximate solution that minimizes both queuing and propagation delay. They further establish that the proposed algorithm gives good performance compared to the traditional Lyapunov method based scheduling algorithm. However, the algorithm derived still sup-optimal and its computational complexity is considerably high.

In [18, 22], the authors try to minimize the average delay in users' queues using Markov Decision Process (MDP) and stochastic learning tools. It has been also adopted in [15] to deal with the problem of power allocation in an OFDM (Orthogonal Frequency Division Multiplexing) system with the goal being to minimize the average delay of the users' packets in the queues. However, the developed solution requires high memory and computational complexity.

### Whittle index for minimizing delay

Whittle index-based policies have been used/developed in wireless networks in the framework of scheduling problems due to its low complexity and its good performance in most cases [23–36]. In [24], a derivation of the Whittle index values for a simple multiclass M/M/1 model has been considered (where only one user can be served). However, the optimality of the obtained Whittle index policy has not been proved. The authors in [37] have considered the problem of project/job scheduling in which an effort is allocated to a fixed number of projects. The performance of a Whittle index-based policy was analyzed under a continuous time model.

Whittle index policy has been derived for a birth-and-death multi-class multi servers queue in [38]. In [39], an optimal index policy called Generalized  $c\mu$ -rule ( $Gc\mu$ ) is developed in the context of heavy-traffic regime with convex delay cost. Furthermore, in contrast to  $c\mu$  rule policy, [40] establishes the optimality of Generalized  $c\mu$ -rule ( $Gc\mu$ ) even with multiple servers.

### 2.4.2 Scheduling Problem: Age Minimization

Among the earliest works that study the AoI minimization problem is that of Kaul, Gruteser, and Yates [4]. They consider a system model where a source is transmitting packets containing status updates to a destination. The channel is represented by a simple queue. In that work, the authors show that neither sending updates as fast as possible nor as low as possible can achieve timeliness goal. Indeed, the optimal solution is not trivial, as one can predict.

In [41], the authors investigate the impact of the waiting room on AoI in a priority-based queuing system where multiple information streams share a common service facility with null or one waiting room. These streams have different priorities. The authors found the optimal arrival rate of status updates that minimizes the average age of each stream.

In scheduling problems, a wide range of works (for instance, [42–48]) have studied these types of problems in order to minimize the age of information in various contexts. To that extent, we cite in the following some interesting works related to our field of studies.

### Energy-aware scheduling

[47] studies a real-time sensing scenario under energy constraints and battery limitations. The authors find an age-optimal threshold policy in function of the energy and the estimated current age states. Another interesting work in this field is [48], where the authors study a scheduling problem for minimizing AoI considering that the source’s update rate cannot exceed a certain predefined limit due to energy limitations. They provide the analytical derivation for the optimal solution when the channel statistics are known.

### Stochastic learning algorithm

Many algorithms have been developed to find the optimal scheduling with respect to AoI based on stochastic learning tools. For instance, [49] proposes an average-cost reinforcement learning algorithm that learns the system parameters, namely channel and energy harvesting statistics, in order to get the optimal policy.

In [50], the authors consider the problem of AoI-optimal scheduling of multiple flows served by a single server. They employ Policy Gradients and Deep Q-Learning methods to achieve scheduling decisions that are resilient to network conditions and packet arrival processes.

On the other hands, a Markov Decision Process (MDP) approximation approach was adopted to propose both off-line and on-line scheduling algorithms in the asymmetric case in [44], which turns out to be asymptotically optimal.

### Whittle index policy for minimizing AoI

As it was indicated in 2.4.1, Whittle index policy has been widely applied when dealing with scheduling problem due to its advantages comparing with other scheduling policies, namely its low complexity and its well performance. Precisely, Whittle index policy is used in Restless Bandit Problem framework which will be detailed in Chapter 3. One of the most known instance of this problem is: How to allocate  $M$  resources to  $N$  users ( $M$  less than  $N$ ) in a such a way to minimize a given metric that evolves stochastically.

[43] tackles this aforementioned problem. Given that RBP is a special case of MDP (Markov decision problem) whose solution is known to be unreachable for many cases, [43] have proved that a greedy algorithm is optimal when the users have identical channel statistics, while for the asymmetric case, they propose as solution Whittle’s index policy which remains sub-optimal. In [45], the authors aim to minimize the total average AoI, under random arrivals by adopting a Whittle’s index-based policy. Accordingly, they derive a simple closed form of the Whittle index policy. They give in addition some numerical results to showcase the performance of the proposed policy. Meanwhile, they don’t prove its optimality. To that extent, we adopt in our work, this well-performing policy, and we further establish its optimality in the asymptotic regime when the number of users scales.

## 2.5 Contributions and overview of the thesis

The remainder of this thesis is structured in three chapters. Specifically:

- Chapter 3: In this chapter, we explain in depth the methodology applied to get the Whittle index policy (WIP). Specifically, we explain in details the steps to get the Whittle indices and we lay out the method used to show the optimality of WIP in the many-users regime.
- Chapter 4: Unlike [24,37–40], we tackle in this Chapter a scheduling problem considering a **discrete time model**. Specifically, we study the delay minimization problem in discrete time queuing system model. To that extent, we investigate a resource allocation problem in which a base station allocates  $M$  channels to  $N$  users' queues at each time slot depending on the scheduling policy considered. The queues evolve according to a Markov process. Furthermore, the users or queues<sup>1</sup> are divided into classes depending on the system parameters namely, packet's arrival rate and packet's transmission rate which differ from one class to another one. Our goal is to find the adequate policy that minimize the long-run total expected average queue length of all users. This aforementioned problem can be cast as Restless Bandit Problem whose solution is known to be unreachable. For that purpose, we apply the low-complex and the well-performing policy, Whittle index policy (WIP) to our specific system model. In a nutshell, we can summarize our contributions into two following points:
  - We derive the Whittle index policy for two scenarios: large buffer size (infinite size), and tight buffer size (less than the transmission rate).
  - We establish the local and the global optimality of WIP in the asymptotic regime when the number of users is very large.
- Chapter 5: In this chapter, we apply Whittle index policy in the context of the age of information (AoI). We consider a scheduling problem with unreliable channels. We consider several classes of users such that each class is characterized by its own channel statistics. Accordingly, our aim is to minimize the long-run total expected average age of all users. Our main contribution in this work, is the proof of asymptotic optimality for both scenarios which are:
  - AoI evolves in finite state's space.
  - AoI evolves in infinite state's space

For the first scenario, we provide the rigorous analysis in order to establish the asymptotic local optimality based on some mathematical techniques. Whilst for the second scenario, we adopt an innovative and original approach to make out the asymptotic global optimality of Whittle index policy. This approach relies essentially on Cauchy criterion: instead of proving that a given function  $f(\cdot)$  converges to a fixed point, we show that the terms of  $f(\cdot)$  are getting closer when  $t$  grows. This method will be detailed in Chapter 5. Finally, we provide some numerical results that bear out our technical analysis.

---

<sup>1</sup>The terms users and queues will be used interchangeably in this Chapter

# Chapter 3

## Methodology

In this thesis, we tackle users and channels scheduling problems related to wireless communication, and we aim to develop a policy called Whittle Index Policy (WIP) that turns out to be optimal under some conditions regarding our metrics of interest, which are **Queuing Delay** and **AoI**. Our problem consists of how to allocate channels to a given subset of users in such a way to optimize the considered metric.

Specifically, from a mathematical point of view, we consider a discrete-time model and an MDP (Markov decision problem) since the system evolves stochastically depending on the system parameters and the policy implemented. Therefore, our purpose will be to minimize the metric of interest that can be viewed as a stochastic objective function, precisely, penalty or cost function that takes discrete values, by adopting the suitable allocation policy. We develop a low complex heuristic policy called Whittle Index Policy, which is optimal when the number of users is large. To that extent, in this chapter, we introduce first the Restless Bandit Problem, then we show that our problem in question is a Restless Bandit Problem (RPB). After that, we give in details the steps to get Whittle's index policy for this kind of problems. Finally, we explain the procedure used to establish the optimality of Whittle's index policy.

### 3.1 Restless Bandit Problem

The Multi-armed problem (MAP) models an agent that attempts to acquire knowledge and share or allocates the available resources among the competing bandits in way to optimize certain value function. The mathematical model can be described as follows: There are  $N$  bandits and only one bandit can be activated at each time slot. A bandit  $k$  is characterized by: the state space  $\mathbf{S}_k$ , an action  $a_k$  that equals to 0 or 1; 0 and 1 refer to the passive and the active action respectively, the transition probability  $p_k(s, s')$  where  $(s, s')$  belongs to  $\mathbf{S}_k \times \mathbf{S}_k$ , and the cost function  $C_k(s, a)$  that depends on the state of bandit  $k$  which is  $s$  as well as the action taken  $a$ . Therefore, the role of a given policy  $\phi$  is to prescribe the active action to one of the bandits while prescribing the passive action to others. We denote by  $\Phi$ , the broad class of scheduling policies that make a scheduling decision based on the history of observed bandit states and scheduling actions. We let  $s_k^\phi(t)$  be the state of bandit  $k$  under policy  $\phi$  at time slot  $t$ . We denote further by  $a_k^\phi(t)$  the action taken with respect to bandit  $k$  under policy  $\phi$ . Given an initial state  $\mathbf{s}(0) = (s_1(0), \dots, s_K(0))$ , the optimal policy  $\phi$  aims to minimize the long-run expected

average cost incurred by the system:

$$\min_{\phi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^N C_k(s_k^\phi(t), a_k^\phi(t)) \mid \mathbf{s}(0) \right] \quad (3.1)$$

Under the following constraint that must be satisfied per each time slot:

$$\sum_{k=1}^N a_k^\phi(t) \leq 1, \text{ for all } t \quad (3.2)$$

MAP has been originally considered by scientists in the second world war, and it has getting more attention after the breakthrough result of Gittin. This later found an optimal policy called Gittin index policy for the MAP problem [52]. However, the shortcoming of the MAP is that  $M=1$  and the bandits for which the action prescribed is the passive action, stay idle, which is not often the case in practice. Thus, Whittle introduces RBP where  $M$  can be higher than one, and all bandits evolve whether the action taken is active or not. That is, the constraint on the available resources per each time slot has a more general expression:

$$\sum_{k=1}^N a_k^\phi(t) \leq M, \text{ for all } t \quad (3.3)$$

## 3.2 System model of interest

In this thesis, as we have already explained in the Introduction, we are interested in minimizing the average delay and the average age. For both cases, we consider  $M$  channels that must be allocated to  $N$  users in such a way to minimize the long-run expected average cost. The action of assigning channel to a given user is the active action. Regarding the first metric, the cost function will be the queue state that evolves stochastically depending on the system's parameters and the action taken. For the second metric AoI, the cost function will be the age of information that also evolves stochastically depending on the system's parameters and the chosen action. After all, we end up with RBP for both metrics.

## 3.3 Lagrangian approach and Whittle index policy

RBPs are PSPACE-Hard (see Papadimitriou et al. [53]), and hence their optimal solution is out of reach. One should, therefore, propose sub-optimal policies when dealing with such problems. In this thesis, we approach the considered RBP problem using the Lagrangian relaxation technique, which consists of relaxing the constraint on the available resources. Instead of having the constraint on the number of available channels satisfied in every time slot, we consider that it has to be met on average. This relaxation allows us to decompose the large relaxed optimization problem into much simpler one-dimensional problems. Based on the optimal solution of the one-dimensional relaxed problems, we develop a heuristic for the original (i.e., non-relaxed) optimization problem. This heuristic is known as the Whittle's index policy (WIP). In this section, we will explain the Lagrangian relaxation method. Then we expose some techniques used to solve RBP based

on the Lagrangian relaxation approach, and we see why these techniques are not feasible in practice. After that, we lay out the steps that allow us to get the Whittle's index Policy.

### 3.3.1 Relaxed Problem and Dual Problem

The Lagrangian relaxation consists of relaxing the constraint of the available resources. Namely, we consider that the constraint in Equation (3.3), has to be satisfied on average and not in every decision epoch, that is,

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{k=1}^N a_k^\phi(t) \right] \leq \alpha N. \quad (3.4)$$

where  $\alpha$  is the proportion of users that are scheduled ( $\alpha N = M$ ). Note that, contrary to the strict constraint in Equation (3.3), the relaxed constraint allows the activation of more than  $\alpha$  fraction of users in each time slot. If we note  $W$  the Lagrangian multiplier for the constrained problem, then the Lagrange function equals to:

$$f(W, \phi) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^N (C_k(s_k^\phi(t), a_k^\phi(t)) + W a_k^\phi(t)) \mid \mathbf{s}(0) \right] - W \alpha N,$$

where  $W$  can be seen as a subsidy for not transmitting. Therefore, the dual problem for a given  $W$  is

$$\min_{\phi \in \Phi} f(W, \phi). \quad (3.5)$$

### 3.3.2 Problem Decomposition

In this section, we show that the relaxed problem can be decomposed into  $N$  one-dimensional subproblems. To do that, we first get rid of the constants that do not depend on  $\phi$  and reformulate the problem as follows,

$$\min_{\phi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^N (C_k(s_k^\phi(t), a_k^\phi(t)) + W a_k^\phi(t)) \mid \mathbf{s}(0) \right]. \quad (3.6)$$

One can see that the solution of this problem is the solution of the well known Bellman equation, see Ross [54], namely,

$$\bar{V}(\mathbf{s}) + \theta = \min_{\mathbf{a}} \left\{ \sum_{k=1}^N C'_k(s_k, a_k) + \sum_{\mathbf{s}'} Pr(\mathbf{s}' \mid \mathbf{s}, \mathbf{a}) \bar{V}(\mathbf{s}') \right\}, \quad (3.7)$$

for all  $\mathbf{s} = (s_1, \dots, s_N)$ , and  $\mathbf{a} = (a_1, \dots, a_N)$ , with  $a_k \in \{0, 1\}$  the action taken with respect to bandit  $k$ . In Equation (3.7),  $V(\cdot)$  represents the Value Function,  $\theta$  is the optimal average cost and  $C'_k(s_k, a_k) = C_k(s_k, a_k) + W a_k$  is the holding cost. The optimal decision for each state  $s$  can be obtained by minimizing the right hand side of Equation (3.7). We now show that the problem can be decomposed into  $N$  independent subproblems by decomposing  $\bar{V}(\cdot)$  into Value Functions for bandit  $k$ , i.e.,  $V_k(\cdot)$ . In other words, the

optimal decision  $\mathbf{a}$  to Problem (3.7) is a vector composed of elements  $a_k$ , where each  $a_k$  is nothing but the optimal decision that solves the one-dimensional Bellman equation:

$$V_k(s_k) + \theta_k = \min_{a_k} \{C'_k(s_k, a_k) + \sum_{s'_k} Pr(s'_k | s_k, a_k) V_k(s'_k)\}. \quad (3.8)$$

This is proven in the next proposition.

**Proposition 3.1.** *Let  $V_k(\cdot)$  be the optimal value function that solves Equation (3.8), and let  $\bar{V}(\cdot)$  be the optimal value function that solves Equation (3.7), then:*

$$\bar{V}(\cdot) = \sum_{k=1}^N V_k(\cdot)$$

*Proof.* See appendix A.1 □

In that sense, resolving (3.8) is equivalent to find an optimal solution of the following problem:

$$\min_{\phi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} (C_k(s_k^\phi(t), a_k^\phi(t)) + W a_k^\phi(t)) \mid s_k(0) \right]. \quad (3.9)$$

Since we deal with the one-dimensional problem, we drop the bandit's index for ease of notation until section 3.4. The solution of Bellman equation (3.8)  $V(\cdot)$ , can be obtained by the well known Value iteration algorithm, which consists in updating  $V_t(\cdot)$  using the following equation:

$$V_{t+1}(s) = \min_a \{C'(s, a) + \sum_{s'} Pr(s' | s, a) V_t(s')\} - \theta \quad (3.10)$$

We consider that the initial value function  $V_0$  equals to 0 for any  $s$ , (i.e. for all  $s$   $V_0(s) = 0$ ). In fact, after many iteration  $V_t(\cdot)$  will converge to the unique fixed point of the equation (3.8) called  $V(\cdot)$  under some conditions (see Puterman [55]). To that extent, the Value iteration algorithm stops when the number of iteration is large enough (when  $V_t(\cdot)$  is nearly constant). We consider that  $t_{inf}$  is the stopping time. As a result, the optimal solution  $\phi^*$  using Value iteration algorithm, verifies for all state  $s$ :

$$\phi^*(s) = \arg \min_a \{C'(s, a) + \sum_{s'} Pr(s' | s, a) V_{t_{inf}}(s')\} \quad (3.11)$$

Dynamic programming solutions and particularly, Value iteration algorithm described above, have been adopted by many existing works that deal with Markov Decision Problem. However these techniques suffer from the curse of dimensionality, which leads to complex resource allocation strategies. In most cases, the developed solutions based on Dynamic Programming require high memory and computational complexity which make them unfeasible. For this reason, we will be limited to give some structural properties of the value function  $V_t(\cdot)$  for any  $t$  to conclude for  $V(\cdot)$ . Based on this, we then establish the structure of the optimal policy of (3.9). Indeed, in our case, after studying the structure of  $V(\cdot)$ , we end up with an optimal solution of type threshold-based policy in function of the state  $s$ :

**Definition 3.1.** An increasing threshold policy is a policy  $\phi \in \Phi$  for which there exists  $n \in \mathcal{S}_k$  such that when the bandit  $k$  is in state  $s \leq n$ , the prescribed action is  $a = 0$ , and when the bandit is in state  $s > n$ , the prescribed action is  $a = 1$ .

In light of that, we derive the indices of Whittle for each state  $s$ , then we express our heuristic policy Whittle Index Policy (WIP). To that end, we provide and explain the steps followed to get the Whittle indices in the next section under the assumption that the optimal solution of Problem (3.8) is threshold-based increasing policy.

### 3.3.3 Whittle's Index

In this section, we provide the derivation of the Whittle indices, which are values that depend on the state  $n$  and system parameters. Although this derivation is made using the relaxed problem, it allows us to develop a heuristic for the original problem. It is worth mentioning that the Whittle's index at a given state, say  $n$ , represents the Lagrange multiplier for which the optimal decision of the one-dimensional dual relaxed problem at this state is indifferent (passive and active decision are both optimal). However, the Whittle index is well defined only if the property of indexability is satisfied. This property requires establishing that as the Lagrange multiplier (or equivalently the subsidy for passivity  $W$ ) increases, the collection of states in which the optimal action is passive increases. In this section, we consider a given bandit  $k$ , and we present two approaches that can be applied to get the Whittle indices. Now, we formalize the indexability and the Whittle's index in the following definitions.

**Definition 3.2.** Considering problem (3.9) for a given  $W$ , we define  $D(W)$  as the set of states in which the optimal action (with respect to the optimal solution of Problem (3.9)) is the passive one. In other words,  $n \in D(W)$  if and only if the optimal action at state  $n$  is the passive one.

$D(W)$  is well defined as the optimal solution of Problem (3.9) is a stationary policy, more precisely, a threshold-based policy by assumption.

**Definition 3.3.** A class is indexable if the set of states in which the passive action is the optimal action increases in  $W$ , that is,  $W' < W \Rightarrow D(W') \subseteq D(W)$ . When the class is indexable, the Whittle's index in state  $n$  is defined as:

$$W(n) = \min\{W | n \in D(W)\} \quad (3.12)$$

In the following, we provide two approaches that can be used to compute the Whittle index values.

#### Discounted Cost Approach

This approach consists in introducing a new expected discounted cost function and deriving the Whittle's index values with respect to the discount parameter  $\beta < 1$ . We then deduce the Whittle's indices for the original problem (i.e., with the total average cost

minimization) by taking the limit  $\beta \rightarrow 1$ . We start by formulating the original problem with the expected discounted cost:

$$\begin{aligned} \min_{\phi \in \Phi} \quad & E \left[ \sum_{t=0}^{+\infty} \sum_{k=1}^N \beta^t C_k(s_k(t), a_k(t)) \mid \mathbf{s}(0), \phi \right], \\ \text{s.t.} \quad & \sum_{k=1}^N a_k(t) \leq \alpha N, \forall t. \end{aligned} \quad (3.13)$$

Following the same steps as in 3.3.1, we relax the problem and give the dual relaxed problem for a given  $W$ :

$$\min_{\phi \in \Phi} \sum_{t=0}^{+\infty} E \left[ \sum_{k=1}^N \beta^t (C_k(s_k(t), a_k(t)) + W a_k(t)) \mid \phi, \mathbf{s}(0) \right]. \quad (3.14)$$

Then, we decompose it into  $N$  one-dimensional problems since the Bellman equation that resolves the dual problem is decomposable. The Bellman equation for an one-dimensional problem is [54].

$$V_\beta(s) = \min_a \{ C'(s, a) + \beta \sum_{s'} Pr(s'|s, a) V_\beta(s') \}. \quad (3.15)$$

In fact,  $V_\beta(s)$  is no more than the discounted cost when the initial state is  $s$ ,  $V_\beta(s) = \sum_{t=0}^{+\infty} E[\beta^t (C'(s(t), a(t))) \mid \phi, s(0) = s]$ .

Unlike the steady-state approach, we can only work with the Bellman equation to derive the Whittle index thanks to the parameter  $\beta$ .

In fact, for most cases, especially for our models, the optimal solution of the Bellman equation (3.15), is of type threshold based policy. On the other hand, under a given threshold policy  $n$ , we are able to give the explicit expression of  $V_\beta(s)$  in function of the system parameters as well as the lagrangian parameter  $W$  for all states  $s$ . According to the definition of Whittle index, we need to check first that the bandit is indexable. This can be verified by showing that the optimal threshold of the problem (3.8) is increasing with subsidy  $W$ . Given that, the Whittle index of state  $s$  denoted by  $W_\beta(s)$  is the langrangian parameter  $W$  that satisfies  $V_\beta^0(s) = V_\beta^1(s)$ , where  $V_\beta^0(s) = C'(s, 0) + \beta \sum_{s'} Pr(s'|s, 0) V_\beta(s')$  and  $V_\beta^1(s) = C'(s, 1) + \beta \sum_{s'} Pr(s'|s, 1) V_\beta(s')$ . Hence, resolving this later equation allows us to find  $W_\beta(s)$  in function of  $s$ ,  $\beta$  and the system parameters. Then under certain conditions on  $V_\beta(\cdot)$  (see [56] and [54]), the Whittle index of state  $s$  under average cost criterion is  $W(s) = \lim_{\beta \rightarrow 1} W_\beta(s)$ .

### Steady State Approach

Let us define  $n$  as the threshold of the bandit  $k$ , i.e., if  $s \leq n$ , then the bandit will not be scheduled, and else, the bandit will be selected. To that extent, the Steady-State Approach consists in the first place of deriving the stationary distribution of the Markov process  $s(t)$  that refers to the state of the bandit  $k$  at time slot  $t$  under the threshold increasing policy  $n$ . Then, we provide a closed-form expression of the Whittle index values based on the steady-state form of the Problem (3.9). Denoting by  $p^n(i, j)$  the transition probability from state  $i$  to  $j$ , by  $u^n$  the stationary distribution under the threshold policy  $n$ , and by  $\mathbf{S}$

the state space of bandit  $k$ , then finding  $u^n$  requires to resolve the full balance equation, i.e.:

$$u^n(i) = \sum_{j \in \mathcal{S}} p^n(j, i) u^n(j) \quad (3.16)$$

With that in mind, we reformulate the dual of the relaxed problem using the stationary distribution. Since the solution of the one-dimensional dual of the relaxed problem (3.9) (given a constant  $W$ ) is a threshold-based policy, we can reformulate the problem as follows:

$$\min_{n \in \mathcal{S}} \mathbb{E}[C^n(s, a) + W a^n] = \min_{n \in \mathcal{S}} \{ \overline{C^n(s, a)} - W \sum_{s=0}^n u^n(s) + W \} \quad (3.17)$$

with  $n$  and  $u^n$  being the threshold and the stationary distribution under the threshold policy  $n$  respectively,  $\overline{C^n(s, a)}$  being the mean cost under threshold policy  $n$

The new formulation of the problem turns out to be useful to derive the Whittle indices since, for any  $W$ , we can find the minimizer of the expression in equation (3.17).

In the literature, several works have been conducted to find Whittle index values. For example, an interesting iterative algorithm has been provided in [57]. In the following, we provide a more general algorithm that can be applied in both cases that we have studied in this thesis (**AoI** and **Queuing Delay**) up to some modifications. In the next chapters, further analysis will be provided to derive a closed-form expression of the Whittle index values based on this algorithm. To that extent, we will first present this algorithm and then prove that it allows the Whittle's index values' computation.

---

**Algorithm 1** Whittle Index Computation

---

- 1: **Init.** Let  $j$  be initialized to 0
  - 2: **Find**  $W_0 = \inf_{n \in \mathbb{N}} \frac{C^n(s, a) - C^{-1}(s, a)}{\sum_{s=0}^n u^n(s)}$
  - 3: **Define**  $n_0$  as the largest minimizer of the above expression
  - 4: **Let**  $W(k) = W_0$  for all  $k \leq n_0$
  - 5: **while**  $n_j \neq \max\{\mathcal{S}\}$  **do**
  - 6:      $j = j + 1$
  - 7:     **Define**  $M_j$  the set  $\{n : \sum_{s=0}^n u^n(s) = \sum_{s=0}^{n_{j-1}} u^{n_{j-1}}(s)\} \cup \{0, \dots, n_{j-1}\}$
  - 8:     **Find**  $W_j = \inf_{n \in \mathbb{N} \setminus M_j} \frac{C^n(s, a) - C^{n_{j-1}}(s, a)}{\sum_{s=0}^n u^n(s) - \sum_{s=0}^{n_{j-1}} u^{n_{j-1}}(s)}$
  - 9:     **Define**  $n_j$  as the largest minimizer of the above expression
  - 10:    **Let**  $W(k) = W_j$  for all  $n_{j-1} < k \leq n_j$
  - 11: **Output** The Whittle index of state  $k$  which is given by  $W(k)$
- 

**Proposition 3.2.** *Assuming that the optimal solution is a threshold policy, and that  $b_n = \sum_{s=0}^n u^n(s)$  is increasing, then the class is indexable. Moreover, if  $a_n = \overline{C^n(s, a)}$  is increasing with  $n$  and for all  $i$  and  $j$  such that  $i < j$   $\sum_{s=0}^i u^i(s) = \sum_{s=0}^j u^j(s) \implies \overline{C^i(s, a)} < \overline{C^j(s, a)}$ , then the Whittle's index values are computed by applying Algorithm 1.*

*Proof.* For the proof, see appendix A.2. □

In chapters 4 and 5, we prove that in our context, the two conditions of Proposition 3.2 are satisfied. Thereby, the algorithm is applicable.

### Whittle's Index Policy

Considering the original Problem (3.8), Whittle's Index Policy, denoted by WIP, will consist simply of allocating the channels to the  $M$  bandits that have the highest Whittle indices at time  $t$  computed using the algorithm 1. As one can remark, Whittle index policy is less complex than the Value iteration algorithm.

## 3.4 Asymptotic Optimality of Whittle Index Policy

The Whittle index policy's performance has been investigated by many works in literature ([30], [32], [23], [58], [34], [35], [29]). This policy turns out to be effectively asymptotically optimal, i.e., it is optimal when the number of bandits and the number of available resources (the number of scheduled bandits per time slot  $t$ ,  $M$ ) scale. In this thesis, we discuss two types of optimality: local and global asymptotic optimality. The Whittle index policy's local optimality means that WIP is asymptotically optimal if the initial state  $s(0)$  belongs to a fixed and restricted set of states. It is often feasible to establish it when the number of bandits is finite. While the global optimality doesn't require any condition of the initial state, it is often hard to be established and needs to take into account some assumptions. In both cases, we show that the long-run average cost under Whittle index policy converges to the optimal cost of the Relaxed problem (the problem (3.1) under the constraint (3.4)) denoted by  $C^{RP,N}$ . The reason behind that is that  $C^{RP,N}$  is a lower bound of all expected average cost obtained by any policy that resolves the original Problem (the problem (3.1) under the constraint (3.3)). Hence, it is sufficient to show that  $\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^N C_k(s_k^\phi(t), a_k^\phi(t)) \mid \mathbf{s}(0)\right]$  converges to  $C^{RP,N}$  under Whittle Index policy when the number of bandits  $N$  and  $T$  grows. In asymmetric cases (the bandits have not the same characteristics), we put together the bandits with the same characteristics in one class. As a consequence, we end up with many classes such that each class contains identical bandits. To that extent, we define  $Z_i^{k,N}$  the proportion of bandits at state  $i$  in class  $k$  over all the system's bandits. In other words, it denotes the number of bandits at state  $i$  in class  $k$  over the number of all bandits, which is  $N$ . If the number of states is finite, then we define the vector  $\mathbf{Z}^N = (Z_1^{1,N}, \dots, Z_{\max \mathbf{s}}^{K,N})$ . In this section, we present briefly the steps followed to prove the asymptotic local and global optimality of Whittle Index Policy when the number of bandit's states is finite.

### 3.4.1 Local Optimality

The local optimality is mainly based on the fluid approximation technique that consists of analyzing the evolution of the expectation of  $\mathbf{Z}^N(t)$  under Whittle's Index policy. For that, we define the vector  $\mathbf{z}(t)$  as follows:

$$\mathbf{z}(t+1) - \mathbf{z}(t)|_{\mathbf{z}(t)=\mathbf{z}} = \mathbb{E}\left[\mathbf{Z}^N(t+1) - \mathbf{Z}^N(t) \mid \mathbf{Z}^N(t) = \mathbf{z}\right] \quad (3.18)$$

In the followings we give the proof outline:

- Showing that  $\mathbf{z}(t)$  converges to the optimal proportion vector of the Relaxed Problem  $\mathbf{z}^*$ , when  $\mathbf{z}(0)$  is within a neighborhood of  $\mathbf{z}^*$ .
- We then connect the fluid approximation model  $\mathbf{z}(t)$  to the discrete-time stochastic system state  $\mathbf{Z}^N(t)$  by using a discrete-time extension of Kurtz's Theorem. Essentially, it states that, over any finite time duration  $[0, T]$ , the actual system evolution  $\mathbf{Z}^N(t)$  can be made arbitrarily close to the above fluid approximation  $\mathbf{z}(t)$  by increasing the number of bandits  $N$ .
- Provided that the initial state is within a neighborhood of  $\mathbf{z}^*$ , we conclude for the convergence of the long-run expected average cost under Whittle index policy to  $C^{RP,N}$  when  $N$  is very large.

### 3.4.2 Global Optimality

We give the proof outline:

- Establishing the existence of a steady-state distribution associated with  $\mathbf{Z}^N(t)$  by demonstrating that  $\mathbf{Z}^N(t)$  under Whittle's index policy evolves in one recurrent class in finite state space.
- Under the assumption that the expected time of reaching a neighborhood of  $\mathbf{z}^*$  doesn't scale with  $N$ , we establish the asymptotic global optimality by adopting the steady-state expression of the long-run expected average cost under Whittle's index policy.

When the number of bandits is infinite, as we have mentioned in Section 2.5, we adopt a new method to prove the global optimality. This method is well explained in Chapter 5.



## Chapter 4

# Whittle's index policy for minimizing the Delay in Queuing systems

In this chapter, we apply the Whittle index policy in a discrete queuing system model. Precisely, we adopt the WIP as solution to the scheduling problem that consists of selecting  $M$  queues among  $N$  queues such that the average queue length of all users is minimized. In the first section, we present our system model. As for the next section, we analyze two scenarios in parallel: The first one is about the case where the buffer size is very small. That is, for each user, the departure rate is greater than the buffer size. While in the second one, the queue length is considered infinite. After that, we express the scheduling problem of interest, then we give the steps of the **Lagrangian Relaxation method** for our particular problem. Then, we prove that the optimal solution of the **Bellman equation** that corresponds to the one-dimensional dual relaxed problem, is effectively a threshold policy. Armed with that, and adopting the **Steady State approach** when the buffer size is finite, we derive Whittle's index policy in closed-form expression in function of the system parameters namely the Buffer length and the departure rate. Whereas, for the infinite buffer size's case, the approach used for the first case when the buffer size is tight fails to give us the expression of Whittle index for all states. To circumvent this difficulty, we handle the problem by applying a new approach named **Discounted Cost approach** in order to get an approximated Whittle index policy for the original problem. Similarly to the first case, we use the **Lagrangian Relaxation technique** and show that the optimal solution of the one-dimensional discounted Bellman equation is a threshold-based policy. Then, we obtain the Whittle index expressions by manipulating the Value function  $V_\beta$  under the discounted parameter  $\beta$ . We conclude for the original problem by tending  $\beta$  to one. Next, we show that Whittle index policy is asymptotically locally optimal using the **fluid approximation technique**, and globally optimal under a recurrence assumption that is verified numerically for our problem. Finally, we give some numerical results that showcase the remarkable good performance of our proposed policy and that corroborate our theoretical findings.

### 4.1 System Model

We consider a time-slotted system with one central scheduler,  $N$  users/queues and  $M$  uncorrelated channels (or servers) with ( $N > M$ ). The terms "server" and "channel" will

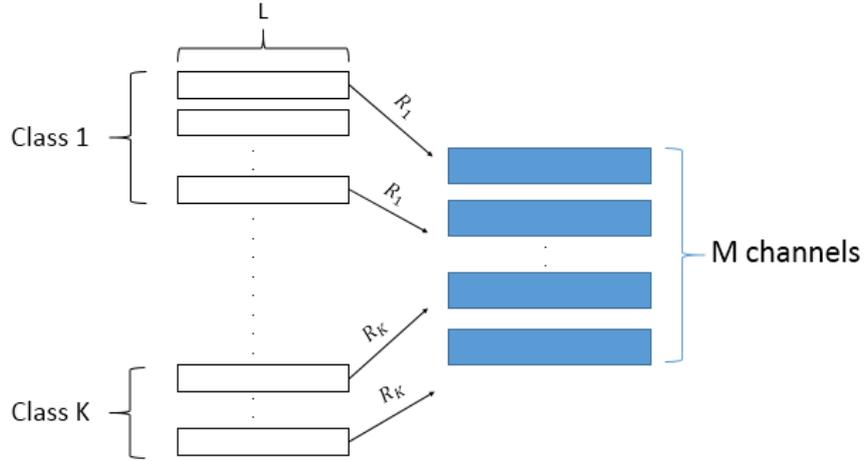


Figure 4.1 – System Model

be used interchangeably throughout this chapter, as well as the terms "user" and "queue". A channel can be allocated to at most one user, hence only  $M$  users will be able to transmit (i.e. send packets) at time slot  $t$ . We consider  $K$  different classes of users and we assume that each user in class- $k$ , if scheduled, transmits at most  $R_k$  packets per time slot. We will refer to  $R_k$  as the maximum transmission rate for every user in class  $k$  and we assume that  $\min_k \{R_k\} \geq 2$ . We denote by  $\gamma_k$  the proportion of class- $k$  users in the system. We further denote by  $A_i^k(t) \in \{0, \dots, R_k - 1\}$  the number of packets that arrive to queue  $i$  in class  $k$  at time slot  $t$  which is independent and identically distributed (i.i.d.) over time. We also let  $q_i^{k,\phi}(t)$  denote the number of packets in queue  $i$  in class  $k$ . Furthermore,  $s_i^{k,\phi}(\mathbf{q}^\phi(t))$  will denote the transmission action under a decision policy  $\phi$  for user  $i$  in class  $k$  and  $\mathbf{q}^\phi(t)$  the vector of all queue lengths  $(q_1^{1,\phi}(t), \dots, q_{N\gamma_1}^{1,\phi}(t), \dots, q_1^{K,\phi}(t), \dots, q_{N\gamma_K}^{K,\phi}(t))$ . For the sake of clarity, we define  $s_i^{k,\phi}(t) := s_i^{k,\phi}(\mathbf{q}^\phi(t))$ . If policy  $\phi$  prescribes to schedule user  $i$  in class  $k$  at time  $t$ , then  $s_i^{k,\phi}(t) = 1$ , and  $s_i^{k,\phi}(t) = 0$  otherwise. We denote by  $L$  the buffer capacity, which is considered to be the same for all queues and can be infinite. The general system model is presented in Figure 4.1. Based on our system model, the number of packets in queue  $i$  of class  $k$  evolves as follows:

$$q_i^{k,\phi}(t+1) = \min\{(q_i^{k,\phi}(t) - R_k s_i^{k,\phi}(t))^+ + A_i^k(t), L\}, \quad (4.1)$$

where  $(x)^+ = \max\{x, 0\}$ .

The objective of the present work is to find a scheduling policy  $\phi$  that minimizes the average queue length of the users which results, according to Little's Law [2] in the minimization of the average delay.

## 4.2 Problem formulation

The cost incurred by user  $i$  in class  $k$ , at time  $t$  is equal to  $a_k q_i^{k,\phi}(t)$  for all  $i \in \{1, \dots, \gamma_k N\}$  where  $a_k$  is a predefined weight. One can see that the model described in Section 4.1 belongs to the family of Restless Bandit Problems (RBP). We consider the broad class  $\Phi$  of scheduling policies in which a scheduling decision depends on the history of observed queue states and scheduling actions. Our user and channel allocation problem therefore

consists of identifying the policy  $\phi \in \Phi$  that minimizes the infinite horizon expected average queues, subject to the constraint on the number of users selected at each time slot. Given the initial state  $\mathbf{q}(0) = (q_1^1(0), \dots, q_{N\gamma_1}^1(0), \dots, q_1^K(0), \dots, q_{N\gamma_K}^K(0))$ , the problem can be formulated as follows:

$$\begin{aligned} & \min_{\phi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} a_k q_i^{k,\phi}(t) \mid \mathbf{q}(0) \right], \\ & \text{s.t. } \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} s_i^{k,\phi}(t) \leq \alpha N, \text{ for all } t, \end{aligned} \quad (4.2)$$

where  $\alpha = M/N$  is the fraction of users that can be scheduled.

### 4.3 Relaxed Problem and Threshold-based Policy

As it has been discussed in Section 3.3, RBPs are PSPACE-Hard and therefore one should develop well performing sub-optimal policies to solve these problems. Accordingly, we have proposed as solution Whittle's index policy for this type of problem. Hence, to derive a closed-form expression of Whittle's index for each state, we apply the methodology explained in Chapter 3. To that extent, in the following, we give briefly the main results for this particular system model.

#### 4.3.1 Relaxed Problem and Dual Problem

The Lagrangian relaxation consists of relaxing the constraint on the available resources. Namely, we consider that the constraint in Equation (4.2), has to be satisfied on average and not in every decision epoch, that is,

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} s_i^{k,\phi}(t) \right] \leq \alpha N. \quad (4.3)$$

If we note  $W$  the Lagrangian multiplier for the constrained problem, then the Lagrange function equals to:

$$f(W, \phi) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} (a_k q_i^{k,\phi}(t) + W s_i^{k,\phi}(t)) \mid \mathbf{q}(0) \right] - W \alpha N, \quad (4.4)$$

where  $W$  can be seen as a subsidy for not transmitting. Therefore, the dual problem for a given  $W$  is

$$\min_{\phi \in \Phi} f(W, \phi). \quad (4.5)$$

#### 4.3.2 Problem Decomposition and Threshold-based Policy

After getting rid of the constants that do not depend on  $\phi$ , the problem (4.5) can be reformulated as follows:

$$\min_{\phi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} (a_k q_i^{k,\phi}(t) + W s_i^{k,\phi}(t)) \mid \mathbf{q}(0) \right]. \quad (4.6)$$

The Bellman equation that corresponds to this problem:

$$V(\mathbf{q}) + \theta = \min_{\mathbf{s}} \left\{ \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} C_k(q_i^k, s_i^k) + \sum_{\mathbf{q}'} Pr(\mathbf{q}'|\mathbf{q}, \mathbf{s}) V(\mathbf{q}') \right\}, \quad (4.7)$$

for all  $\mathbf{q} = (q_1^1, \dots, q_{\gamma_1 N}^1, \dots, q_1^K, \dots, q_{\gamma_K N}^K)$ , with  $q_i^k \in \{1, \dots, L\}$  being the queue length of class- $k$  user  $i$ , and  $\mathbf{s} = (s_1^1, \dots, s_{\gamma_1 N}^1, \dots, s_1^K, \dots, s_{\gamma_K N}^K)$ , with  $s_i^k \in \{0, 1\}$  being the action taken with respect to user  $i$  in class  $k$ . In equation (4.7),  $V(\cdot)$  represents the Value Function,  $\theta$  is the optimal average cost and  $C_k(q_i^k, s_i^k)$  is the holding cost  $a_k q_i^k + W s_i^k$ . According to the Chapter 3, section 3.3.2, the problem can be decomposed into  $N$  independent sub-problems, one for each user  $i$  in class  $k$ . Accordingly, resolving the Bellman equation (4.7) is equivalent to resolve the following Bellman equation for each user  $i$  in class  $k$ .

$$V_i^k(q_i^k) + \theta_i^k = \min_{s_i^k} \left\{ C_k(q_i^k, s_i^k) + \sum_{q_i'^k} Pr(q_i'^k | q_i^k, s_i^k) V_i^k(q_i'^k) \right\}. \quad (4.8)$$

In the following, for both cases of buffer size  $L$  ( $L < R_k$  for all  $k$ , or  $L$  is infinite), we show that the solution to each one-dimensional problem (for each user  $i$ ) follows the structure of a threshold policy. For ease of notation, we drop the indices  $k$  and  $i$  and consider that  $V(\cdot)$  is the value function for a given user. From this perspective, we let  $A$  be the random variable that indicates the number of arrival packets at each time slot.

As it has been indicated in Chapter 3 Section 3.3.2, the solution of the Bellman equation (4.8),  $V(\cdot)$  can be obtained by the well known Value iteration algorithm:

$$V_{t+1}(q) = \min_s \left\{ C(q, s) + \sum_{q'} Pr(q'|q, s) V_t(q') \right\} - \theta \quad (4.9)$$

Providing that  $\lim_{t \rightarrow +\infty} V_t(\cdot) = V(\cdot)$ , we give some structural properties of the value function  $V_t(\cdot)$  for any  $t$  and conclude for  $V(\cdot)$ . Then we give the structure of the optimal solution. To that end, we distinguish between two cases of  $L$ :

**Remark 4.1.** *It is worth to emphasize that when  $L$  is less than  $R$ , if the arrival packets plus the current queue length overflow on the buffer capacity, the user retain only the  $L$  packets and get rid of the surplus of the packets. Subsequently, from the state  $q$ , we can reach the state  $L$  when the number of arrival packets  $A$  can be either  $L - q$  or plus. Having said that, nor  $Pr(L|q, 1) = Pr(A = L - (q - R)^+) = Pr(A = L)$  neither  $Pr(L|q, 0) = Pr(A = L - q)$ , rather  $Pr(L|q, 1) = \sum_{A=L}^{R-1} Pr(A = L)$  and  $Pr(L|q, 0) = \sum_{A=L-q}^{R-1} Pr(A = L - q)$ . While if  $L$  is infinite, since there is no overflow on the buffer capacity, for all queue states  $q$  and  $q'$ , we have that  $Pr(q'|q, 1) = Pr(A = q' - (q - R)^+)$  and  $Pr(q'|q, 0) = Pr(A = q' - q)$ .*

$L < R$

To tackle this case, we proceed with these following steps:

- We prove that  $V(\cdot)$  is increasing with  $q$

- We establish that  $V^1(q) - V^0(q)$  is decreasing with  $q$  where  $V^0(q)$  and  $V^1(q)$  are the value functions when the action prescribed at state  $q$  is  $s = 0$  and  $s = 1$  respectively
- Finally, we show that the optimal solution is an increasing threshold policy

Regarding the first point, we show that  $V_t(\cdot)$  is increasing with  $q$  for all  $t$  by induction. Precisely, we establish that:

- $V_0(\cdot)$  increases with  $q$ .
- If  $V_t(\cdot)$  is increasing with  $q$ ,  $V_{t+1}(\cdot)$  is also increasing with  $q$ .

Given that  $V_0(\cdot) = 0$ , then  $V_0(\cdot)$  is increasing with  $q$ .

Considering  $V_t(\cdot)$  is increasing with  $q$ , then  $\sum_{q'} Pr(q'|\cdot, s)V_t(q')$  grows with  $q$  (see Puterman [55]). We have by construction,  $C(\cdot, s)$  increases with  $q$ . Since  $\theta$  is just a constant,  $V_{t+1}(\cdot)$  will be as well an increasing function with  $q$ .

As consequence, we show that  $V_t(\cdot)$  is increasing with  $q$  for all  $t$ . Leveraging the fact that  $V(\cdot)$  is the limit of  $V_t(\cdot)$  when  $t$  grows, then  $V(\cdot)$  is also increasing with  $q$ .

As for the second point, one can see that for any  $q$ , we have that  $q \leq L < R$ . Then the next state before the arrival of the packets will be  $q = 0$  if the action prescribed is the active action since all packet in the buffer will be transmitted. Consequently, the probability to transit to the state  $q'$  from a given state  $q$  under the active action is the probability to have  $A = q'$  if  $q' < L$ , or  $L \leq A \leq R - 1$  if  $q' = L$  (according to the remark 4.1). Hence  $Pr(q'|q, 1)$  doesn't depend on  $q$ . Likewise for  $\sum_{q'=0}^L Pr(q'|q, 1)V(q')$ .

We have that:

$$V^1(q) - V^0(q) = W + \sum_{q'} Pr(q'|q, 1)V(q') - \sum_{q'} Pr(q'|q, 0)V(q')$$

Bearing in mind that  $V(\cdot)$  is increasing with  $q$ , then again according to Puterman [55],  $\sum_{q'} Pr(q'|q, 0)V(q')$  is increasing with  $q$ . Leveraging the above result,  $\sum_{q'=0}^L Pr(q'|q, 1)V(q')$  is constant with respect to  $q$ . Consequently,  $V^1(q) - V^0(q)$  decreases with  $q$ .

To prove the last point, we recall that the optimal action  $s(q)$  at state  $q$  according to Equation (4.7), is the one that minimizes  $V^s(q)$ . Explicitly,  $s(q) = \operatorname{argmin}\{V^0(q), V^1(q)\}$ . Moreover, exploiting the fact that  $V^1(q) - V^0(q)$  is decreasing with  $q$ , then there exists  $q_0$  (can be infinite) such that for all  $q \leq q_0$ ,  $V^1(q) \geq V^0(q)$  and for all  $q > q_0$ ,  $V^1(q) \leq V^0(q)$ . Consequently, we deduce that for all  $q \leq q_0$ , the optimal decision is to stay idle, and for all  $q > q_0$ , the optimal decision is to transmit. Thereby, we prove that the optimal solution of Problem (4.5) is of type threshold increasing policy.

$L = +\infty$

We consider the operator  $TO$  such that for each  $(q, s) \in \mathbb{N} \times \{0, 1\}$

$$(TO(V))(q, s) = C(q, s) + \sum_{q'} Pr(q'|q, s)V(q') - \theta \quad (4.10)$$

We first provide some useful definitions and preliminary results before proving the desired results.

**Definition 4.1.** We say that a function  $f$  is  $R$ -convex in  $X = \mathbb{N}$ , if for any  $x$  and  $y$  in  $X$  such that  $x < y$ , we have:

$$f(y + R) - f(x + R) \geq f(y) - f(x) \quad (4.11)$$

**Lemma 4.1.** If for a given function  $f$ , there exists  $R$  such that for any integer  $x$ ,  $f(x + 1 + R) - f(x + R) \geq f(x + 1) - f(x)$ , then  $f$  is  $R$ -convex

*Proof.* Considering  $y$  and  $x$  in  $\mathbb{N}$ , with  $y > x$ , we have:

$$f(y + R) - f(x + R) = \sum_{k=x}^{y-1} [f(k + 1 + R) - f(k + R)] \quad (4.12)$$

$$\geq \sum_{k=x}^{y-1} [f(k + 1) - f(k)] \quad (4.13)$$

$$= f(y) - f(x) \quad (4.14)$$

which concludes the proof.  $\square$

**Definition 4.2.** Let  $g(q, s)$  be a real valued function defined on  $X \times S$ , with  $S = \{0, 1\}$ , and  $X = \mathbb{N}$ . We say that  $g$  is submodular if  $g(q + 1, 1) - g(q + 1, 0) \leq g(q, 1) - g(q, 0)$  for all  $q$  on  $X$ .

**Theorem 4.1.**  $\min_s TO(\cdot)(\cdot, s)$  conserves the  $R$ -convexity and the increase properties. In other words, if the input of the operator  $TO$ , i.e. a given  $V(\cdot)$ , is  $R$ -convex and increasing function with  $q$  then  $\min_s TO(V)(\cdot, s)$  is  $R$ -convex and increasing function with  $q$ .

*Proof.* Let us consider that the input of  $TO(V)(\cdot, s)$ , i.e. a given  $V(\cdot)$ , is  $R$ -convex and increasing function with  $q$ . For the increase property of  $\min_s TO(V)(\cdot, s)$ , we have by definition that  $C(\cdot, s)$  is increasing with  $q$ . We also have that  $V(\cdot)$  is increasing with  $q$ , then  $\sum_{q'} Pr(q'|\cdot, s)V(q')$  grows with  $q$  (see Puterman [55]). Since  $\theta$  is a constant,  $TO(V)(\cdot, s)$  increases with  $q$  and therefore  $\min_s TO(V)(\cdot, s)$  is increasing with  $q$ .

For  $R$ -convexity, we should first prove the following lemma.

**Lemma 4.2.** If  $V(\cdot)$  is  $R$ -convex and increasing with  $q$ ,  $C(q, s)$  and  $\sum_{q'} Pr(q'|q, s)V(q')$  are submodular functions.

*Proof.* The proof is given in appendix B.1.  $\blacksquare$

This demonstrates that the function  $TO(V)(\cdot, \cdot)$  is submodular since it is the sum of two submodular functions. Let us now show that  $\min_s TO(V)(\cdot, s)$  is  $R$ -convex. For that, we consider the function  $\Delta TO(V)(q) = TO(V)(q, 1) - TO(V)(q, 0)$  which is decreasing with  $q$  since  $TO(V)(\cdot, \cdot)$  is submodular. Therefore, there exists  $r \in \mathbb{R} \cup \{+\infty\}$  such that for  $q \leq r$ ,  $\Delta TO(V)(q) \geq 0$  and for  $q \geq r$ ,  $\Delta TO(V)(q) \leq 0$ . In the remainder of the proof, we consider all possible cases of  $q$  and  $r$ .

If  $q + R + 1, q + R, q, q + 1 \geq r$ :

$$\min_s TO(V)(q + 1 + R, s) - \min_s TO(V)(q + R, s) = TO(V)(q + 1 + R, 1) - TO(V)(q + R, 1) \quad (4.15)$$

$$= TO(V)(q + 1, 0) - TO(V)(q, 0) \quad (4.16)$$

$$\geq TO(V)(q + 1, 1) - TO(V)(q, 1) \quad (4.17)$$

$$= \min_s TO(V)(q + 1, s) - \min_s TO(V)(q, s) \quad (4.18)$$

where the inequality is due to the sub-modularity of  $TO(V)(\cdot, \cdot)$ .

If  $q \leq r \leq q + 1, q + R, q + 1 + R$ :

$$\min_s TO(V)(q + 1 + R, s) - \min_s TO(V)(q + R, s) = TO(V)(q + 1 + R, 1) - TO(V)(q + R, 1) \quad (4.19)$$

$$= TO(V)(q + 1, 0) - TO(V)(q, 0) \quad (4.20)$$

$$\geq TO(V)(q + 1, 1) - TO(V)(q, 1) \quad (4.21)$$

$$= \min_s TO(V)(q + 1, s) - \min_s TO(V)(q, s) \quad (4.22)$$

if  $q, q + 1 \leq r \leq q + R, q + R + 1$ :

$$\min_s TO(V)(q + 1 + R, s) - \min_s TO(V)(q + R, s) = TO(V)(q + 1 + R, 1) - TO(V)(q + R, 1) \quad (4.23)$$

$$= TO(V)(q + 1, 0) - TO(V)(q, 0) \quad (4.24)$$

$$= \min_s TO(V)(q + 1, s) - \min_s TO(V)(q, s) \quad (4.25)$$

if  $q, q + 1, q + R \leq r \leq q + R + 1$ :

$$\min_s TO(V)(q+1+R, s) - \min_s TO(V)(q+R, s) = TO(V)(q+1+R, 1) - TO(V)(q+R, 0) \quad (4.26)$$

$$\geq TO(V)(q+1+R, 1) - TO(V)(q+R, 1) \quad (4.27)$$

$$= TO(V)(q+1, 0) - TO(V)(q, 0) \quad (4.28)$$

$$= \min_s TO(V)(q+1, s) - \min_s TO(V)(q, s) \quad (4.29)$$

If  $q, q+1, q+R; q+R+1 \leq r$ :

$$\min_s TO(V)(q+1+R, s) - \min_s TO(V)(q+R, s) = TO(V)(q+1+R, 0) - TO(V)(q+R, 0) \quad (4.30)$$

$$\geq TO(V)(q+1+R, 1) - TO(V)(q+R, 1) \quad (4.31)$$

$$= TO(V)(q+1, 0) - TO(V)(q, 0) \quad (4.32)$$

$$= \min_s TO(V)(q+1, s) - \min_s TO(V)(q, s) \quad (4.33)$$

Using Lemma 4.1,  $\min_s TO(V)(\cdot, s)$  is R-convex with  $q$ , i.e., we can conclude the R-convexity conservation.

□

**Remark 4.2.** Theorem 1 means that if the value function  $V_t$  is increasing and R-convex, then the value function  $V_{t+1}$  in equation (4.9), which is computed with the operator  $TO$ , is increasing and R-convex.

Thus, as  $V_0$  is increasing and R-convex, all  $V_t$  are increasing and R-convex and therefore we can conclude that the value function  $V$  will be also R-convex and increasing with  $q$ .

**Corollary 4.1.** The optimal policy  $\phi^*$  of each one-dimensional relaxed subproblem is a threshold-based policy.

*Proof.* Since we have only two possible actions, a policy is of the form threshold policy if and only if it is monotone with  $q$ . Therefore, it is sufficient to prove that the optimal policy  $\phi^*$  is monotone with  $q$ . For that, we consider  $q_1 \leq q_2$ . According to Remark 4.2,  $V(\cdot)$  is increasing and R-convex, then using Lemma 4.2,  $TO(V)$  is submodular. Therefore, we have:

$$(TO(V))(q_1, 1) - (TO(V))(q_1, 0) \geq (TO(V))(q_2, 1) - (TO(V))(q_2, 0) \quad (4.34)$$

If  $\phi^*(q_2) = \underset{s}{\operatorname{argmin}}(TO(V))(q_2, s) = 0$

Hence,  $(TO(V))(q_2, 1) - (TO(V))(q_2, 0) \geq 0$ . Consequently, we get:

$$(TO(V))(q_1, 1) - (TO(V))(q_1, 0) \geq 0 \quad (4.35)$$

Which leads to:

$$\underset{s}{\operatorname{argmin}}(TO(V))(q_1, s) = 0 \quad (4.36)$$

i.e.

$$\phi^*(q_1) \leq \phi^*(q_2) \quad (4.37)$$

If  $\phi^*(q_2) = \underset{s}{\operatorname{argmin}}(TO(V))(q_2, s) = 1$ , obviously we have that:

$$\phi^*(q_1) \leq \phi^*(q_2) \quad (4.38)$$

Therefore, we can conclude that the optimal solution is monotone and increasing with  $q$ , then it is a threshold-based policy.  $\square$

## 4.4 Whittle's Index

For  $L < R$ , we apply the Steady State approach described in Chapter 3, Section 3.3.3 to find the Whittle's indices.

### 4.4.1 Steady State approach ( $L < R$ )

Let us define  $n$  as the threshold for users in a given class, i.e. if the queue state is  $q$  such that  $q \leq n$ , then the user will not be scheduled, and else, the user will be selected for transmission. The objective of this section is to derive the stationary distribution of the users' states. We assume here that the packets arrive according to a discrete uniform distribution, that is,  $\mathbb{P}(A(t) = x) = \rho$  for all  $0 \leq x \leq R - 1$  and 0 otherwise, where  $\rho = 1/R$ .

We denote by  $p^n(i, j)$  the transition probability from state  $i$  to  $j$ , by  $u$  the stationary distribution under the threshold policy  $n$ , and by  $R$  the maximum rate ( $\rho = 1/R$ ). One can notice that  $u$  verifies the full balance equation, i.e.:

$$u(i) = \sum_{j=0}^L p^n(j, i)u(j) = \sum_{j=0}^n p^n(j, i)u(j) + \sum_{j=n+1}^L p^n(j, i)u(j) \quad (4.39)$$

**Definition 4.3.** We define  $\pi_i$  as:

$$\pi_i = \begin{cases} \rho & \text{if } 0 \leq i \leq R - 1 \\ 0 & \text{else} \end{cases} \quad (4.40)$$

**Proposition 4.1.** The expressions of  $p^n(j, i)$  are given by:  
If  $0 \leq i < L$  and  $j \leq n$

$$p^n(j, i) = \pi_{i-j} = \begin{cases} \rho & \text{if } 0 \leq i - j \leq R - 1 \\ 0 & \text{else} \end{cases} \quad (4.41)$$

if  $0 \leq i < L$  and  $n < j \leq L$

$$p^n(j, i) = \pi_i = \begin{cases} \rho & \text{if } 0 \leq i \leq R - 1 \\ 0 & \text{else} \end{cases} \quad (4.42)$$

if  $i = L$  and  $j \leq n$

$$p^n(j, L) = (R - L + j)\pi_{L-j} = (R - L + j)\rho \quad (4.43)$$

if  $i = L$  and  $n < j \leq L$

$$p^n(j, L) = (R - L)\pi_L = (R - L)\rho \quad (4.44)$$

*Proof.* See appendix B.2. □

**Proposition 4.2.** *The expressions of the stationary distribution is:*

$$u(i) = \begin{cases} \rho(1 - \rho)^{n-i} & \text{if } 0 \leq i \leq n \\ \rho & \text{if } n + 1 \leq i \leq L - 1 \\ (1 - \rho)^{n+1} - (L - n - 1)\rho & \text{if } i = L \end{cases} \quad (4.45)$$

2)  $n = L$ :

$$u(i) = \begin{cases} 0 & \text{if } 0 \leq i \leq L - 1 \\ 1 & \text{if } i = L \end{cases} \quad (4.46)$$

*Proof.* See appendix B.3. □

We reformulate the dual of the relaxed problem using the stationary distribution derived above, i.e.,

$$\min_{n \in [-1, L]} \mathbb{E}[aq^n + Ws^n] = \min_{n \in [-1, L]} \left\{ \sum_{q=0}^L au^n(q)q - W \sum_{q=0}^n u^n(q) + W \right\} \quad (4.47)$$

with  $n$  and  $u^n$  being the threshold and the stationary distribution under the threshold policy  $n$ .

We first give the expression of the mean cost in equation (4.47) given threshold  $n$  when  $L < R$ .

If  $-1 \leq n \leq L - 1$ :

$$\sum_{q=0}^L au^n(q)q = a \left[ (L + R)(1 - \rho)^{n+1} + n - R + 1 + \frac{(L - 1 - n)(n - L)}{2R} \right] \quad (4.48)$$

If  $n = L$ :

$$\sum_{q=0}^L au^n(q)q = aL \quad (4.49)$$

Secondly, we provide the expression of the passive decision's average time in equation (4.47) given a threshold  $n$ .

If  $-1 \leq n \leq L - 1$ :

$$\sum_{q=0}^n u^n(q) = 1 - (1 - \rho)^{n+1} \quad (4.50)$$

If  $n = L$ :

$$\sum_{q=0}^n u^n(q) = 1 \quad (4.51)$$

According to Chapter 3 Section 3.3.3, to get the values of Whittle's indices, we apply the algorithm 1. Nevertheless, we need to adapt it to this case before using it. To that extent, the adapted algorithm will be:

---

**Algorithm 2** Whittle Index Computation

---

- 1: **Init.** Let  $j$  be initialized to 0
  - 2: **Find**  $W_0 = \inf_{n \in \mathbb{N}} \frac{\sum_{q=0}^L au^n(q)q - \sum_{q=0}^L au^{-1}(q)q}{\sum_{q=0}^n u^n(q)}$
  - 3: **Define**  $n_0$  as the largest minimizer of the above expression
  - 4: **Let**  $W(k) = W_0$  for all  $k \leq n_0$
  - 5: **while**  $n_j \neq L$  **do**
  - 6:      $j = j + 1$
  - 7:     **Define**  $M_j$  the set  $\{0, \dots, n_{j-1}\}$
  - 8:     **Find**  $W_j = \inf_{n \in \mathbb{N} \setminus M_j} \frac{\sum_{q=0}^L au^n(q)q - \sum_{q=0}^L au^{n_{j-1}}(q)}{\sum_{q=0}^n u^n(q) - \sum_{q=0}^{n_{j-1}} u^{n_{j-1}}(q)}$
  - 9:     **Define**  $n_j$  as the largest minimizer of the above expression
  - 10:    **Let**  $W(k) = W_j$  for all  $n_{j-1} < k \leq n_j$
  - 11: **Output** The Whittle index of state  $k$  which is given by  $W(k)$
- 

**Remark 4.3.** In order to simplify the notation in the sequel, we denote  $\sum_{q=0}^L au^n(q)q$  by  $a_n$  and  $\sum_{q=0}^n u^n(q)$  by  $b_n$ .

However, in order to apply Algorithm 2 that allows to obtain the Whittle's index for each state, we need to check that the conditions given in Proposition 3.2 are satisfied. We start first by establishing the indexability.

**Theorem 4.2.** For each  $k$ , the class- $k$  is indexable.

*Proof.* According to Proposition 3.2, we just need to prove that  $\sum_{q=0}^n u^n(q)$  is increasing with  $n$ . It is clear that from (4.50),  $\sum_{q=0}^n u^n(q)$  is increasing with  $n$ . Hence, the class is indexable.  $\square$

We prove the two others conditions of Proposition 3.2 which are the increase property of  $\sum_{q=0}^L au^n(q)q$  with  $n$ , and that for all  $i$  and  $j$  such that  $i < j$ ,  $\sum_{q=0}^i u^i(q) = \sum_{q=0}^j u^j(q) \implies \sum_{q=0}^L au^i(q)q < \sum_{q=0}^L au^j(q)q$ . The second property for this case is meaningless since  $\sum_{q=0}^i u^i(q)$  is strictly increasing with  $i$ . While for the first one, one should demonstrate that  $a_n$  is increasing with  $n$ .

**Proposition 4.3.**  $a_n = \sum_{q=0}^L au^n(q)q$  is increasing with  $n$ .

*Proof.* See appendix B.4. □

As the indexability is satisfied and the two conditions of Proposition 3.2 are verified, then we can apply Algorithm 2 to get the Whittle's index for each state. However, the complexity of this algorithm is  $L^2$ . In order to overcome this complexity issue, we will provide further analysis and derive simple expressions of the Whittle indices. We first proceed by laying out the following definitions and lemmas.

**Definition 4.4.** For any given increasing threshold policy  $n$ , we define  $y^n$  as a function of the subsidy  $W$ , such that  $y^n(W) = \sum_{q=0}^L au^n(q)q - W \sum_{q=0}^n u^n(q) = a_n - Wb_n$ .

**Lemma 4.3.** The intersection point  $W = x_{i,j}$  between  $y^i(W)$  and  $y^j(W)$  is equal to:

$$x_{i,j} = \frac{\sum_{q=0}^L au^i(q)q - \sum_{q=0}^L au^j(q)}{\sum_{q=0}^i u^i(q) - \sum_{q=0}^j u^j(q)} \quad (4.52)$$

**Theorem 4.3.** The Whittle index of state  $n \in [0, L]$ :

$$W(n) = x_{n,n-1} = \frac{a[\rho[(L-n) - (L+R)(1-\rho)^n] + 1]}{\rho(1-\rho)^n}$$

*Proof.* See appendix B.5 □

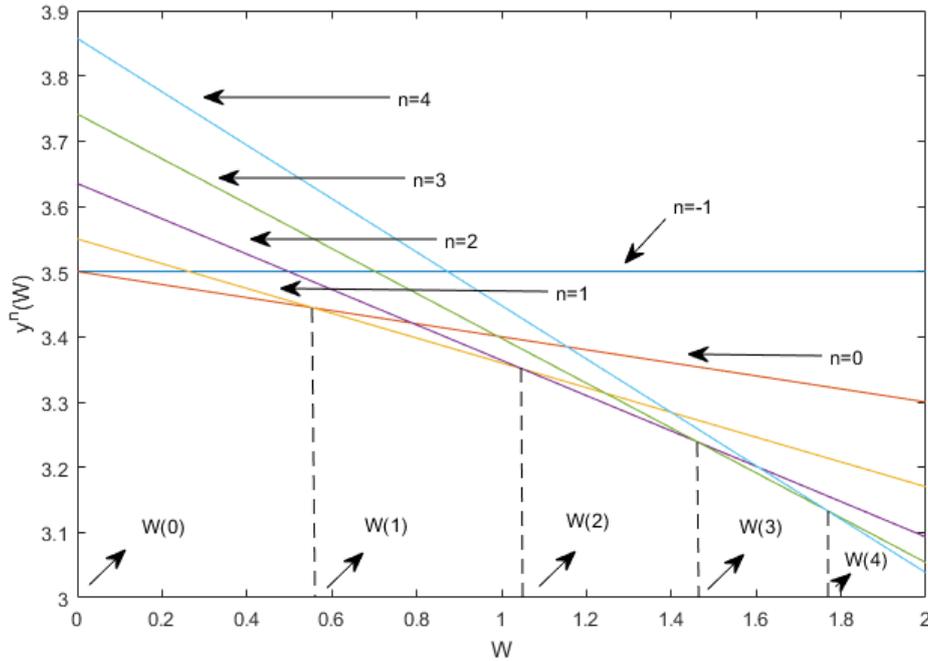


Figure 4.2 –  $y^n(W)$  in function of  $W$  for different values of  $n$  ( $L = 5$  and  $R = 10$ ):  $W(i)$  indicates the Whittle index at state  $i$

### Whittle's Index policy for the original problem

We now consider the original optimization problem (4.2) and propose a simple Whittle's Index policy. This policy consists of simply allocating the channels to the  $M$  users that have the highest Whittle indices at time  $t$ , denoted by  $WIP$ , and computed using the simple expressions in Theorem 4.3.

Now, we tackle the second case: when  $L$  is infinite. For that, we use the Discounted Cost approach detailed in Chapter 3 Section 3.3.3.

#### 4.4.2 Discounted Cost approach ( $L = +\infty$ )

In this section, we will adopt the Discounted Cost approach to get the Whittle indices. We will also explain the limitation of the steady state approach, and why it cannot be used to find the Whittle index values if the queues have an unlimited capacity.

Applying the steady state approach in the unlimited capacity framework, we get the following results:

- Computing the stationary distribution under a given threshold policy  $n$ , we end up with an increasing average passive time with  $n$ , i.e.,  $\sum_{i=0}^n u^n(i)$  grows with  $n$ . Therefore, according to Proposition 3.2, the problem is indexable.
- The Whittle's indices expressions are defined only for states in  $[0, R_k - 1]$  and are given by:

$$\text{For } 0 \leq n \leq R_k - 1: W_k(n) = \frac{a_k R_k n}{R_k - n}$$

As one can see, the above result is limited to the case where the states are in  $[0, R_k - 1]$ . For this reason, we rely on another method which allows us to find the Whittle index values for all possible states. To do so, we formulate a discounted cost problem in which  $\beta$  is a discount factor. We analyze this discounted problem and find the Whittle index expressions (that depends on  $\beta$ ). Then by taking  $\beta \rightarrow 1$ , we obtain the Whittle index for our original problem. Similarly, since the Discounted Cost approach is already explained in Chapter 3 Section 3.3.3, we give briefly the main theoretical findings.

The original problem with the expected discounted cost is:

$$\begin{aligned} \min_{\phi \in \Phi} \quad & E \left[ \sum_{t=0}^{+\infty} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} \beta^t a_k q_i^k(t) \mid q(0), \phi \right], \\ \text{s.t.} \quad & \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} s_i^k(t) \leq \alpha N, \forall t. \end{aligned} \tag{4.53}$$

The dual of the relaxed problem for a given  $W$  is:

$$\min_{\phi \in \Phi} \sum_{t=0}^{+\infty} E \left[ \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} \beta^t (a_k q_i^k(t) + W s_i^k(t)) \mid \phi, q(0) \right]. \tag{4.54}$$

The Bellman equation for an one-dimensional problem is:

$$V(q_i^k) = \min_{s_i^k} \{ C(q_i^k, s_i^k) + \beta \sum_{q_i'^k} Pr(q_i'^k \mid q_i^k, s_i^k) V(q_i'^k) \}. \tag{4.55}$$

Following the same method in 4.3.2 but with a parameter  $\beta$  instead of  $\theta$ , we can prove that the optimal solution that satisfies this Bellman equation is a threshold policy, by proving that the function  $TO(V)$  is submodular. We can also conclude that the value function has the same structural property as in Section 4.3.2, especially, the submodularity and  $R_k$ -convexity. However, contrary to the method applying when  $L$  is finite, finding the steady-state distribution will not give an explicit expression of Problem (4.54). Nevertheless, we can work only with the Bellman equation to derive the Whittle index thanks to the parameter  $\beta$  which helps us to find the Whittle index for all states.

From now on, we detail the procedure followed to obtain the Whittle index policy for the discounted Cost problem (4.54).

To prove that the Whittle index for a given state  $n$  in class  $k$  is a given  $W_k(n)$ , according to the definition of the Whittle's index, we have to demonstrate that for all  $W \leq W_k(n)$ , at state  $n$  the optimal decision of the Bellman equation (4.55) must be the active action. And for all  $W > W_k(n)$ , at state  $n$  the optimal decision of (4.55) must be the passive action. In other terms, since the optimal solution of (4.55) is a threshold-based policy, it is sufficient to show that for all  $W \leq W_k(n)$ , the optimal threshold policy must be strictly less than  $n$ , and for all  $W > W_k(n)$ , the optimal threshold policy must be greater than  $n$ . Accordingly, if we can find a function that links a given state  $n$  to the corresponding lagrangian parameter  $W$  for which  $n$  is the optimal threshold policy with respect to the problem (4.55), then we will be able to characterize the Whittle's index of state  $n$  by investigating the variation of this function with respect to  $n$ . To that end, we proceed in two steps: finding this aforementioned function, and studying its evolution. Regarding the first step, we start by giving these two following definitions.

**Definition 4.5.** We define  $C_0^n(q_i^k, W)$  and  $C_1^n(q_i^k, W)$  in class  $k$  as the discounted costs starting at the initial queue state  $q_i^k$  at which the decision taken is to not be scheduled ( $s_i^k = 0$ ) or to be scheduled ( $s_i^k = 1$ ) respectively and when the policy considered is threshold  $n$ , explicitly:

$$C_0^n(q_i^k, W) \triangleq a_k q_i^k + \beta \sum_{q_i'^k} Pr(q_i'^k | q_i^k, 0) V^n(q_i'^k, W),$$

$$C_1^n(q_i^k, W) \triangleq a_k q_i^k + W + \beta \sum_{q_i'^k} Pr(q_i'^k | q_i^k, 1) V^n(q_i'^k, W),$$

where  $V^n(\cdot, W)$  is the value function under threshold policy  $n$  and lagrangian parameter  $W$ .

**Definition 4.6.** We define  $g_k(n, W)$  as a function defined in  $[0, +\infty[ \times \mathbb{R}$ , such that for all  $(n, W) \in [0, +\infty[ \times \mathbb{R}$ ,  $g_k(n, W) = C_1^n(n, W) - C_0^n(n, W)$

**Proposition 4.4.** For  $0 \leq n \leq R_k - 1$ :

$$g_k(n, W) = W(1 - n\beta\rho_k) - a_k n\beta$$

For  $n \geq R_k$ :

$$g_k(n, W) = \frac{W(1 - \beta) - a_k R_k \beta}{1 - \rho_k \beta}$$

*Proof.* See Appendix B.6. □

We emphasize that to prove that for a fixed  $W$ , in class  $k$ , a given state  $n$  is indeed an optimal threshold (i.e. if  $q_i^k \leq n$  the queue is not scheduled and otherwise it is scheduled), we just need to prove that it satisfies for all states  $q_i^k \leq n$ ,  $C_0^n(q_i^k, W) \leq C_1^n(q_i^k, W)$  and for  $q_i^k > n$ ,  $C_0^n(q_i^k, W) \geq C_1^n(q_i^k, W)$  (according to Bellman equation). In other words, we suppose that  $n$  is a threshold (i.e. if  $q_i^k \leq n$  the queue is not scheduled and otherwise it is scheduled), and we show that for all states  $q_i^k \leq n$ ,  $C_0^n(q_i^k, W) \leq C_1^n(q_i^k, W)$  and for  $q_i^k > n$ ,  $C_0^n(q_i^k, W) \geq C_1^n(q_i^k, W)$ . We note that for a given value of  $W$ , the optimal threshold might not be unique.

**Proposition 4.5.** *For a given lagrangian parameter  $W$  and class  $k$ , if there exists  $n$  such that  $C_0^n(n, W) = C_1^n(n, W)$ , then  $n$  is an optimal threshold.*

*Proof.* See Appendix B.7. □

According to the proposition above, the lagrangian parameter  $W$  for which  $n$  is the optimal threshold is the one that verifies,  $g_k(n, W) = 0$ . Subsequently, as we have already derived the explicit expression of  $g_k(\cdot, \cdot)$ , we can find the expression of the parameter  $W$  in question in function of  $n$ . For that purpose, we provide this following proposition.

**Proposition 4.6.** *If  $n \leq R_k - 1$ , then for  $W = \frac{\beta a_k R_k n}{R_k - \beta n}$ ,  $n$  is an optimal threshold.*

*If  $n \geq R_k$ , then for  $W = \frac{a_k R_k \beta}{1 - \beta}$ ,  $n$  is an optimal threshold.*

*Proof.* See Appendix B.8. □

As for the second step, we study the function  $g_k$  defined in Definition 4.6.

**Lemma 4.4.**  *$g_k$  is strictly increasing with  $W$ , and decreasing with  $n$ .*

*Proof.*  $g_k$  is clearly strictly increasing with  $W$  and decreasing with  $n$  from its expression. □

Leveraging the above result, we lay out our desired result in this following Theorem.

**Theorem 4.4.** *For each queue state  $n$  in class  $k$ , the Whittle index expression is given by:*

$$\text{For } 0 \leq n \leq R_k - 1, W_k(n) = \frac{\beta a_k R_k n}{R_k - \beta n}.$$

$$\text{For } n \geq R_k, W_k(n) = \frac{a_k R_k \beta}{1 - \beta}.$$

*Proof.* See Appendix B.9. □

We know that for  $\beta \rightarrow 1$ , the solution for Problem (4.54) is the same as Problem (4.6), see [54]. Hence, to derive the Whittle index for the expected average cost's case, we must compute the limit as  $\beta$  tends to 1. However, for states greater or equal than  $R_k$ , the Whittle indices tend to  $+\infty$ . On the other hand, by looking at our policy which consists on selecting the users at states with the  $M$  highest Whittle's index values, we can notice that this policy is the same if the order of the Whittle indices from the biggest to the smallest one is not affected even if the Whittle index values are modified.

**Theorem 4.5.** *For any  $\beta > 1 - \frac{\min\{a_j R_j\}}{\max\{a_j R_j^2\}}$ , the Whittle index policy where the Whittle indices in each class  $k$  are given by:*

$$\text{For } 0 \leq n < R_k: W_k(n) = \frac{\beta a_k R_k n}{R_k - \beta n}$$

$$\text{For } R_k \leq n: W_k(n) = a_k R_k \max\{a_j R_j^2\}$$

*is exactly the Whittle index policy when the Whittle indices are given by Theorem 4.4.*

*Proof.* See Appendix B.10. □

### Whittle's Index policy for the original problem

When  $\beta \rightarrow 1$ , the condition given in Theorem 4.5 still true. Hence, we get the Whittle index policy for our original problem with the expected average cost given by Theorem 4.5. We can notice that the Whittle index of states greater than the maximum transmission rate are different only by  $a_k$  and  $R_k$  ( $W_k(n) = a_k R_k \max\{a_j R_j^2\}$ ). Thereby, it is worth mentioning that the obtained policy can be seen as  $c\mu$  rule when all states are greater than  $R_k$ , since we choose  $M$  users with the highest  $a_k R_k$ .

## 4.5 Further analysis of the optimal solution of the relaxed problem

As it has been mentioned in Section 3.4, to establish the asymptotic optimality, we compare the average cost under Whittle's index policy with the optimal cost of the relaxed problem denoted  $C^{RP,N}$ . For this reason, we should provide further analysis and give the structure of the optimal solution for the relaxed problem. From now on, we focus only on the case where  $L$  is finite. As we have seen in Section 4.3, for any given  $W$ , the optimal solution for the dual relaxed problem (4.6) is a threshold-based policy for each user. By using the Whittle index expressions derived in Theorem 4.3, we provide a derivation of the optimal threshold for each class as a function of the Lagrange parameter  $W$ . In this section, we denote by  $W_i^k$  the Whittle index at state  $i$  in class  $k$ . We denote by  $l = (l_1, l_2, \dots, l_K)$  the vector which represents the set of thresholds for each class  $k$ . We denote by  $u_k^n$ , the stationary distribution for class  $k$  under threshold policy  $n$ .

**Proposition 4.7.** *For a given  $W$ , the optimal threshold vector  $l = (l_1(W), l_2(W), \dots, l_K(W))$  for the dual problem satisfies:*

For each  $k$ :

$$l_k(W) = \arg \max_i \{W_i^k | W_i^k \leq W\} \quad (4.56)$$

or

$$l_k(W) = \arg \max_i \{W_i^k | W_i^k < W\} \quad (4.57)$$

We note that the solution can also be a linear combination between the threshold policies  $\arg \max_i \{W_i^k | W_i^k \leq W\}$  and  $\arg \max_i \{W_i^k | W_i^k < W\}$ .

*Proof.* See appendix B.11. □

Now, we give the structure of the optimal solution of the constrained relaxed problem.

**Proposition 4.8.** *The solution of the constrained relaxed problem is of type threshold policy  $l(W^*)$ , with  $l$  being the function vector defined in Proposition 4.7 and  $W^*$  satisfies  $\alpha = \sum_{k=1}^K \gamma_k \sum_{i=l_k(W^*)+1}^L u_k^{l_k(W^*)}(i)$ .*

*Proof.* See appendix B.12. □

However,  $W^*$  that satisfies the above constraint may not exist since  $\alpha$  is a real number that can take any value in  $[0, 1]$ , and  $\sum_{k=1}^K \gamma_k \sum_{i=l_k(W)+1}^L u_k^{l_k(W)}(i)$  is discrete, since the vector  $l(W)$  can only take discrete values in  $[-1, L]^K$ . To deal with this issue, we use the

fact that for some values of  $W$ , the optimal solution of the dual problem can be a linear combination or more precisely a randomized policy between two threshold policies for a given class as it has been mentioned in Proposition 4.7. To that extent, our task is to find among these values of  $W$ , the one for which there exists a randomized parameter  $\theta$  such that the constraint is satisfied with equality. To that end, we introduce this following proposition.

**Proposition 4.9.** *There exists a class  $m$ , state  $p$ , and a randomization parameter  $\theta^*$  such that the optimal solution of the dual problem when the langrangian parameter  $W = W_p^m = W^*$  is characterized by:*

- For  $k \neq m$ , the optimal threshold is  $l_k(W_p^m) = \arg \max_i \{W_i^k | W_i^k \leq W_p^m\}$
- For  $k = m$ , the optimal solution is randomized policy between two threshold policies  $l_m(W_p^m) = \arg \max_i \{W_i^m | W_i^m \leq W_p^m\}$  and  $l_m(W_p^m) - 1 = \arg \max_i \{W_i^m | W_i^m < W_p^m\}$ , where the factor of randomization  $\theta^*$  is the probability of adopting the policy  $l_m(W_p^m)$  and  $1 - \theta^*$ , the probability of adopting the policy  $l_m(W_p^m) - 1$ .
- The constraint (4.3) is satisfied with equality, i.e.

$$\alpha = \sum_{k \neq m} \sum_{i=l_k(W_p^m)+1}^L \gamma_k u_k^{l_k(W_p^m)}(i) + \sum_{i=l_m(W_p^m)+1}^L \gamma_m u_m^*(i) + (1 - \theta^*) \gamma_m u_m^{l_m(W_p^m)-1}(l_m(W_p^m))$$

$$\text{Where } u_m^* = \theta^* u_m^{l_m(W_p^m)} + (1 - \theta^*) u_m^{l_m(W_p^m)-1}.$$

*Proof.* See appendix B.13 □

The solution of the dual problem described in Proposition 4.9 satisfies the constraint (4.3) with equality, then according to Proposition 4.8, this solution is indeed the optimal solution of the constrained problem. In that regard, the optimal cost of the relaxed problem  $C^{RP,N}$ , is expressed as following:

$$C^{RP,N} = \sum_{k \neq m} \sum_{i=0}^L N \gamma_k a_k u_k^{l_k(W_p^m)}(i) i + \sum_{i=0}^L N \gamma_m a_m u_m^*(i) i \quad (4.58)$$

## 4.6 Local asymptotic optimality ( $L < R$ )

In this section, we will show that the performance of the Whittle's Index policy is asymptotically locally optimal when  $L$  is finite. As we have clarified in Chapter 3 Section 3.4, we will compare the average cost obtained by the Whittle's Index policy; WIP; with the one obtained for the relaxed problem RP to establish the optimality of WIP. Explicitly, denoting by  $C_T^N(\mathbf{x})$  the average cost obtained over the time duration  $0 \leq t \leq T$  under Whittle's Index policy conditioned on the initial state  $\mathbf{x}$ , we show that  $C_T^N(\mathbf{x})$  tends to  $C^{RP,N}$  when  $N$  and  $T$  scale. For that, we will be in need of the relaxed problem's optimal cost's expression,  $C^{RP,N}$  derived in Section 4.5.

We reconsider the notion of the proportion defined in Chapter 3 Section 3.4 that represents the number of queues at state  $i$  in class  $k$  over the number of all users which is  $N$ .

We have that  $\mathbf{Z}^N = (\mathbf{Z}^{1,N}, \dots, \mathbf{Z}^{K,N})$  with  $\mathbf{Z}^{k,N} = (Z_1^{k,N}, \dots, Z_L^{k,N})$  and  $\sum_{i=0}^L Z_i^{k,N} = \gamma_k$  for each class  $k$ .

The expression of  $C_T^N(\mathbf{x})$  in function of  $\mathbf{Z}^N$  is  $\frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k Z_i^{k,N}(t) i N \mid \mathbf{Z}^N(0) = \mathbf{x} \right]$ ,

where  $\mathbf{Z}^N(t)$  evolves under Whittle's Index policy. Denoting by  $\mathbf{z}^*$  the optimal proportion of the relaxed problem, we say that the Whittle's Index policy is asymptotically locally optimal if there exists  $\delta > 0$  such that the initial proportion vector  $\mathbf{Z}^N(0)$  is within  $\Omega_\delta(\mathbf{z}^*)$  (i.e.  $\|\mathbf{Z}^N(0) - \mathbf{z}^*\| < \delta$ ), then  $C_T^N(\mathbf{x})$  converges to  $C^{RP,N}$  when  $T$  and  $N$  scale.

In order to prove that, we use the fluid limit technique introduced in Chapter 3, i.e., we investigate the evolution of the expectation of  $\mathbf{Z}^N(t)$  under the Whittle's Index policy. The expectation of  $\mathbf{Z}^N(t)$  denoted by  $\mathbf{z}(t)$ , verifies:

$$\mathbf{z}(t+1) - \mathbf{z}(t) |_{\mathbf{z}(t)=\mathbf{z}} = \mathbb{E} [\mathbf{Z}^N(t+1) - \mathbf{Z}^N(t) | \mathbf{Z}^N(t) = \mathbf{z}] \quad (4.59)$$

If we denote by  $w_j^h$  the Whittle index for class  $h$  at state  $j$  and by  $p_i^k(\mathbf{z})$  the probability that a user is selected randomly among  $z_i^k$  to transmit, one can easily show that [37]:

$$p_i^k(\mathbf{z}) = \min\{z_i^k, \max(0, \alpha - \sum_{w_j^h > w_i^k} z_j^h)\} / z_i^k \quad (4.60)$$

We denote by  $q_{i,j}^{k,0}$  and  $q_{i,j}^{k,1}$  the probabilities of transition from state  $i$  to state  $j$  in a class- $k$  if the queue is not scheduled and if the queue is scheduled for transmission respectively. Then, the probability of transition from state  $i$  to state  $j$  in class  $k$  is:

$$q_{i,j}^k(\mathbf{z}) = p_i^k(\mathbf{z}) q_{i,j}^{k,1} + (1 - p_i^k(\mathbf{z})) q_{i,j}^{k,0} \quad (4.61)$$

Accordingly, we have that for each  $i$  and  $k$ :

$$z_i^k(t+1) - z_i^k(t) = \sum_{j \neq i} q_{j,i}^k(\mathbf{z}(t)) z_j^k(t) - \sum_{i \neq j} q_{i,j}^k(\mathbf{z}(t)) z_i^k(t) \quad (4.62)$$

Let  $w^*$  be the Lagrangian parameter that gives the optimal solution of the relaxed problem. Then, according to Proposition 4.9, there exists a given class  $m$  such that  $w_{l_m}^m = w^*$  for which the corresponding optimal solution of the relaxed problem is of type threshold policy for class  $k \neq m$  denoted  $l_k$ , and a randomized policy between two threshold policies  $l_m$  and  $l_m - 1$  for class  $m$ .

We define  $J_{w^*}$  as the set of states such that at any system state  $\mathbf{z} \in J_{w^*}$ , if we use the Whittle's Index policy, all users with the Whittle index value higher than  $w^*$  are scheduled, while the users with Whittle index value smaller than  $w^*$  stay idle and the users with Whittle index value  $w^*$  are scheduled with a certain randomization. Specifically,  $J_{w^*} = \{\mathbf{z} : \sum_{w_i^k > w^*} z_i^k < \alpha, \sum_{w_i^k \geq w^*} z_i^k \geq \alpha\}$ .

Providing that for all  $k$  and  $t$ :

$$\sum_{j=0}^L z_j^k(t) = \gamma_k \quad (4.63)$$

Therefore, the following equation always holds for  $\mathbf{z}(t) \in J_{w^*}$ :

1)  $k \neq m$ :

$$z_i^k(t+1) = \sum_{j=0}^{l_k-1} (q_{j,i}^{k,0} - q_{l_k,i}^{k,0}) z_j^k(t) + \sum_{j=l_k+1}^L (q_{j,i}^{k,1} - q_{l_k,i}^{k,0}) z_j^k(t) + \gamma_k q_{l_k,i}^{k,0} \quad (4.64)$$

2)  $k = m$

$$\begin{aligned}
 z_i^m(t+1) &= \sum_{j=0}^{l_m-1} (q_{j,i}^{m,0} - q_{l_m,i}^{m,0}) z_j^m(t) + \sum_{j=l_m+1}^L (q_{j,i}^{m,1} - q_{l_m,i}^{m,1}) z_j^m(t) + (1-\alpha) q_{l_m,i}^{m,0} + \alpha q_{l_m,i}^{m,1} \\
 &\quad - \left( \sum_{\substack{w_j^h > w_{l_m}^m \\ h \neq m, j \neq l_h}} z_j^h(t) q_{l_m,i}^{m,1} - \left( \sum_{\substack{w_j^h \leq w_{l_m}^m \\ h \neq m, j \neq l_h}} z_j^h(t) q_{l_m,i}^{m,0} + \left( \sum_{\substack{h=1 \\ h \neq m}}^K \sum_{\substack{j=0 \\ j \neq l_h}}^L \mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} z_j^h(t) \right) q_{l_m,i}^{m,1} \right. \right. \\
 &\quad \left. \left. + \left( \sum_{\substack{h=1 \\ h \neq m}}^K \sum_{\substack{j=0 \\ j \neq l_h}}^L \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} z_j^h(t) \right) q_{l_m,i}^{m,0} - \sum_{\substack{h=1 \\ h \neq m}}^K \gamma_h \left( \mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} q_{l_m,i}^{m,1} + \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} q_{l_m,i}^{m,0} \right) \right) \right)
 \end{aligned} \tag{4.65}$$

Let  $g_i^m = \sum_{\substack{h=1 \\ h \neq m}}^K \gamma_h (\mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} q_{l_m,i}^{m,1} + \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} q_{l_m,i}^{m,0}) \forall i \in [0, L]$ , and  $C = (c^1, \dots, c^K)$  such that  $c^k = (\gamma_k q_{l_k,0}^{k,0}, \dots, \gamma_k q_{l_k,L}^{k,0})$  and  $c^m = ((1-\alpha) q_{l_m,0}^{m,0} + \alpha q_{l_m,0}^{m,1} - g_0^m, \dots, (1-\alpha) q_{l_m,L}^{m,0} + \alpha q_{l_m,L}^{m,1} - g_L^m)$  for each  $k \neq m$ .

Then,

1)  $k \neq m$ :

$$z_i^k(t+1) = \sum_{j=0}^{l_k-1} (q_{j,i}^{k,0} - q_{l_k,i}^{k,0}) z_j^k(t) + \sum_{j=l_k+1}^L (q_{j,i}^{k,1} - q_{l_k,i}^{k,1}) z_j^k(t) + c_i^k \tag{4.66}$$

2)  $k = m$ :

$$\begin{aligned}
 z_i^m(t+1) &= \sum_{j=0}^{l_m-1} (q_{j,i}^{m,0} - q_{l_m,i}^{m,0}) z_j^m(t) + \sum_{j=l_m+1}^L (q_{j,i}^{m,1} - q_{l_m,i}^{m,1}) z_j^m(t) + c_i^m \\
 &\quad - \left( \sum_{\substack{w_j^h > w_{l_m}^m \\ h \neq m, j \neq l_h}} z_j^h(t) q_{l_m,i}^{m,1} - \left( \sum_{\substack{w_j^h \leq w_{l_m}^m \\ h \neq m, j \neq l_h}} z_j^h(t) q_{l_m,i}^{m,0} + \left( \sum_{\substack{h=1 \\ h \neq m}}^K \sum_{\substack{j=0 \\ j \neq l_h}}^L \mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} z_j^h(t) \right) q_{l_m,i}^{m,1} \right. \right. \\
 &\quad \left. \left. + \left( \sum_{\substack{h=1 \\ h \neq m}}^K \sum_{\substack{j=0 \\ j \neq l_h}}^L \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} z_j^h(t) \right) q_{l_m,i}^{m,0} \right) \right)
 \end{aligned} \tag{4.67}$$

Then, by replacing in the equation above for all  $k$   $z_{l_k}^k(t)$  with  $\gamma_k - \sum_{j=0, j \neq l_k}^L z_j^k(t)$ , we obtain the following linear relation in  $J_{w^*}$  between  $\tilde{z}(t+1)$  and  $\tilde{z}(t)$  where  $\tilde{z}$  is the proportion vector in which the elements  $z_{l_k}^k$  for different  $k$  are eliminated.

$$\tilde{z}(t+1) = \mathbf{Q} \tilde{z}(t) + \mathbf{C} \tag{4.68}$$

where  $\mathbf{C}$  is a constant matrix and the expression of matrix  $\mathbf{Q}$  is given in Appendix B.14. The vector solution of the relaxed problem, denoted by  $\tilde{z}^*$ , is the fixed point of the aforementioned linear equation. By definition of  $\tilde{z}^*$ ,  $\tilde{z}^* \in J_{w^*}$ . Thus, if  $\tilde{z}(0) = \tilde{z}^* + \mathbf{e}$  and  $\tilde{z}(t) \in J_{w^*}$ , we obtain for  $t$ :

$$\tilde{z}(t) - \tilde{z}^* = \mathbf{Q}^t \mathbf{e} \tag{4.69}$$

The analysis of the above linear system is therefore important to prove the local optimality. We first provide the following lemma.

**Lemma 4.5.** *If for all eigenvalues  $\lambda$  of  $\mathbf{Q}$ ,  $|\lambda| < 1$ , then there exists a neighborhood  $\Omega_\sigma(\tilde{\mathbf{z}}^*) \subseteq \mathcal{J}_{w^*}$  such that if  $\tilde{\mathbf{z}}(0) \in \Omega_\sigma(\tilde{\mathbf{z}}^*)$ , we have the following:*

- 1) For all  $t \geq 0$ ,  $\|\tilde{\mathbf{z}}(t) - \tilde{\mathbf{z}}^*\| < \sigma$  ( $\tilde{\mathbf{z}}(t) \in \mathcal{J}_{w^*}$ ).
- 2)  $\tilde{\mathbf{z}}(t)$  converges to  $\tilde{\mathbf{z}}^*$ .

*Proof.* The proof follows from the convergence of the linear system. □

**Proposition 4.10.** *For all eigenvalue  $\lambda$  of  $\mathbf{Q}$ ,  $|\lambda| < 1$*

*Proof.* See the proof in appendix B.14. □

The aforementioned result, combined with Lemma 4.5, proves the convergence of the fluid limit system. Consequently,  $\mathbf{z}(t)$  converges to the fixed point of Equation (4.59),  $\mathbf{z}^*$ . However, the above result is not enough to prove the local optimality, as we have to show that the stochastic vector  $\mathbf{Z}^N(t)$  converges to  $\mathbf{z}^*$  in probability. For that, we introduce the discrete-time version of Kurtz Theorem applied to our problem (see [59]):

**Proposition 4.11.** *There exists a neighborhood  $\Omega_\delta(\mathbf{z}^*)$  of  $\mathbf{z}^*$  such that if  $\mathbf{Z}^N(0) = \mathbf{z}(0) = \mathbf{x} \in \Omega_\delta(\mathbf{z}^*)$ , then for any  $\mu > 0$  and finite time horizon  $T$ , there exist positive constants  $C_1$  and  $C_2$  such that*

$$P_{\mathbf{x}}(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu) \leq C_1 \exp(-NC_2) \quad (4.70)$$

where  $\delta < \sigma$ , and  $P_{\mathbf{x}}$  denotes the probability conditioned on the initial state  $\mathbf{Z}^N(0) = \mathbf{z}(0) = \mathbf{x}$ . Furthermore,  $C_1$  and  $C_2$  are independent of  $\mathbf{x}$  and  $N$ .

According to the above proposition, the system state  $\mathbf{Z}^N(t)$  behaves very closely to the fluid approximation model  $\mathbf{z}(t)$  when the number of users  $N$  is large. Since we have shown the convergence of  $\mathbf{z}(t)$  to within  $\Omega_\sigma(\mathbf{z}^*)$ , we are ready to establish the local convergence of the system state  $\mathbf{Z}^N(t)$  to  $\mathbf{z}^*$ .

**Lemma 4.6.** *If  $\mathbf{Z}^N(0) = \mathbf{x} \in \Omega_\delta(\mathbf{z}^*)$ , then for any  $\mu > 0$ , there exists a time  $T_0$  such that for any  $T > T_0$ , there exists positive constants  $s_1$  and  $s_2$  with,*

$$P_{\mathbf{x}}(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu) \leq s_1 \exp(-Ns_2) \quad (4.71)$$

*Proof.* See appendix B.15. □

Now we are ready to prove the asymptotic local optimality of the proposed scheduling policy.

**Proposition 4.12.** *If the initial state is in the set  $\Omega_\delta(\mathbf{z}^*)$ , then*

$$\lim_{T \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} = \frac{C^{RP,N}}{N} \quad (4.72)$$

*Proof.* See appendix B.16 □

## 4.7 Global asymptotic optimality ( $L < R$ )

In this section, we prove that from any initial state  $\mathbf{x}$ , the long-run expected average cost obtained with the Whittle's Index policy is optimal when  $N$  is very large. In contrast to the method used to prove the local optimality, we work here with the steady state distribution of the stochastic process  $\mathbf{Z}^N(t)$ . To ensure that such a stationary distribution exists, we need to show that there is at least one recurrent state. Since the states evolve according to a finite state Markov chain, we just need to prove that there exists a state reachable from any other states.

**Lemma 4.7.** *The state defined as for each class  $k$ ,  $\mathbf{z}^k = (1, 0, \dots, 0)$  and denoted by  $\mathbf{z}_0^1$ , is reachable from any initial state using the Whittle's Index policy.*

*Proof.* See appendix B.17 □

This lemma is stronger than proving the existence of a recurrent state. Indeed, this allows us to deduce that  $\mathbf{Z}^N(t)$  evolves in one recurrent aperiodic class, and that there exists a stationary distribution for  $\mathbf{Z}^N(t)$  denoted by  $\mathbf{Z}^N(\infty)$ . We still need to check if for a fixed  $N$ , there exists at least one recurrent state within  $\Omega_\epsilon(\mathbf{z}^*)$ , as otherwise  $\Omega_\epsilon(\mathbf{z}^*)$  will be a transient class. If such a state exists, surely  $\mathbf{Z}^N(t)$  will evolve in one recurrent class that contains this recurrent state. To that end, we demonstrate here that  $\mathbf{z}^*$  is reachable from any state for a fixed  $N$ . Since  $\mathbf{z}_0$  is reachable from any state, we just need to find a path from  $\mathbf{z}_0$  to  $\mathbf{z}^*$ .

**Lemma 4.8.** *By applying the Whittle's Index policy, the steady state  $\mathbf{z}^*$  is reachable from the state  $\mathbf{z}_0$ .*

*Proof.* Since  $L < R$ , the probability of transitioning from queue state 0 to any other state whether the action is active or passive is strictly positive. Thereby, there exists a trajectory from  $\mathbf{z}_0$  to  $\mathbf{z}^*$  which lasts only one time slot. Consequently,  $\mathbf{z}^*$  is reachable from  $\mathbf{z}_0$ . □

From this lemma above, the state  $\mathbf{z}^*$  is reachable from any state, which means that  $\mathbf{z}^*$  is a recurrent state. However, the considered actions schedule a proportion of users (i.e. not an integer value). This is not feasible and unrealistic for some (small) values of  $N$  since the queues are not splittable. In fact, for some values of  $N$ , the state  $\mathbf{z}^*$  may not exist. On the other hand, we can say that for enough large  $N$ , and for any  $\epsilon > 0$ , there exists at least one recurrent state within the neighborhood  $\Omega_\epsilon(\mathbf{z}^*)$ . This will ensure that there is a path to enter a neighborhood  $\Omega_\epsilon(\mathbf{z}^*)$  from any initial state. However, it is important to ensure that the time to enter  $\Omega_\epsilon(\mathbf{z}^*)$  should not scale up with  $N$ . For that, we give the following assumption which will be later justified via numerical studies in Section 4.8.

**Assumption 4.1.** *We assume that the expected time to enter a neighborhood of  $\mathbf{z}^*$  from any initial state  $\mathbf{x}$  does not depend on the number of queues  $N$ . In other words, for all  $N$  the time to enter a neighborhood  $\Omega_\epsilon(\mathbf{z}^*)$  denoted by  $\Gamma_{\mathbf{x}}^N(\epsilon)$  is bounded by a constant  $T_{b_\epsilon}$ .*

Now we provide a useful lemma that allows us to demonstrate the global asymptotic optimality.

---

<sup>1</sup> $\mathbf{z}_0$  can be seen as the system's state where all queues are in the queue state 0

**Lemma 4.9.** *Under assumption 4.1, and for any  $\epsilon$ , we have that:*

$$\lim_{N \rightarrow +\infty} P(\mathbf{Z}^N(\infty) \in \Omega_\epsilon(\mathbf{z}^*)) = 1 \quad (4.73)$$

*Proof.* See Lemma 6 in [32]. □

Since we have found a stationary distribution of  $\mathbf{Z}^N(t)$  under Whittle's Index policy, the expected average cost under Whittle's Index policy for a fixed  $N$  can be written as follows:

$$\lim_{T \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} = \sum_{k=1}^K \sum_{i=0}^L a_k \mathbb{E} \left[ Z_i^{k,N}(\infty) \right] iN \quad (4.74)$$

**Theorem 4.6.** *Under assumption 4.1, and for any initial state, we have that:*

$$\lim_{N \rightarrow +\infty} \lim_{T \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} = \frac{C^{RP,N}}{N} \quad (4.75)$$

*Proof.* See appendix B.18 □

## 4.8 Numerical Results

In this section, we provide numerical results that confirms the asymptotic optimality of the developed Whittle index policy. To that extent, we study two scenarios depending on the buffer size  $L$ :

- $L \neq +\infty$
- $L = +\infty$

### 4.8.1 Numerical Results for finite buffer size

We consider 2 classes having a respective rate of  $R_1 = 15$  and  $R_2 = 20$ . Moreover, we suppose that  $\alpha = 1/2$ ,  $L = 10$ ,  $\gamma_1 = \gamma_2 = 1/2$ , and  $a_1 = a_2 = a = 1$ . We also consider two initial states  $x$  and  $y$  such that all the queues are equal to 0 and  $L$  respectively.

#### Verification of Assumption 4.1

We plot in Figure 4.3, the evolution of the time needed to enter a neighborhood  $\Omega_\epsilon(\mathbf{z}^*)$  (i.e. hitting time of  $\Omega_\epsilon(\mathbf{z}^*)$ ) with respect to  $N$ , given that  $\epsilon$  is small enough.

One can see that for large values of  $N$ , the hitting time can be considered as a constant and does not diverge for both initial states  $\mathbf{x}$  and  $\mathbf{y}$ . This implies that the hitting time is bounded for large values of  $N$  which consolidates Assumption 4.1.

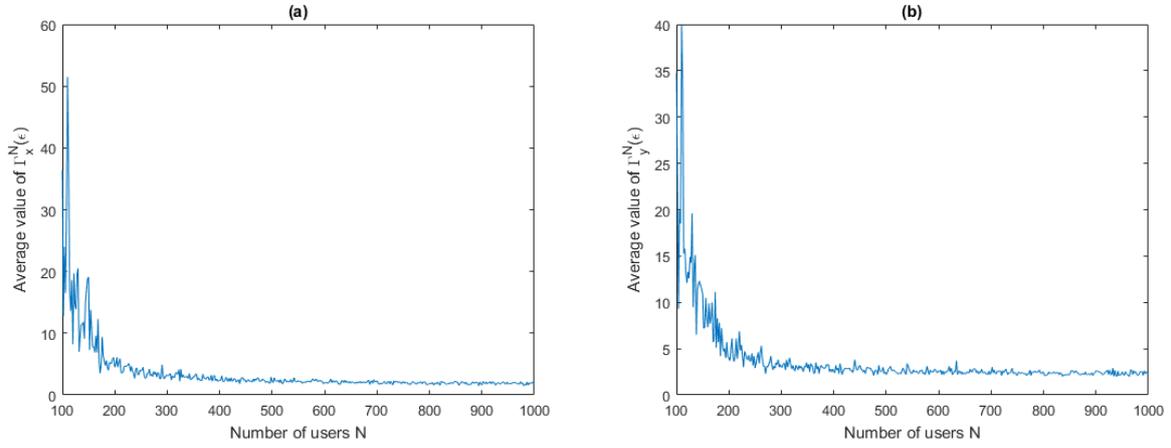


Figure 4.3 – Hitting Time of  $\Omega_\epsilon(\mathbf{z}^*)$  in function of  $N$ : (a)  $\mathbf{Z}^N(0) = \mathbf{x}$ , (b)  $\mathbf{Z}^N(0) = \mathbf{y}$

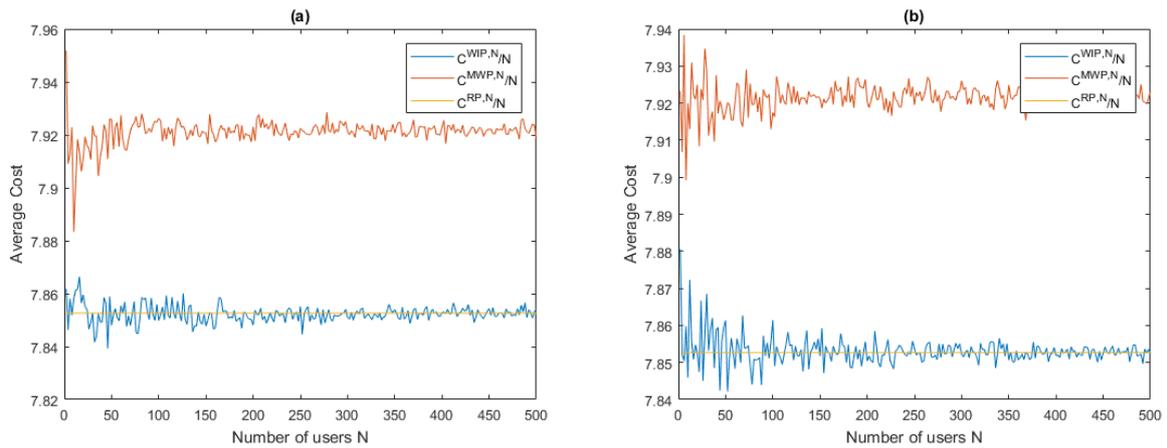


Figure 4.4 – Performance evaluation of Whittle's Index policy

## Performance of the Whittle's Index policy

In this section, we compare the long-run expected average cost per user under the Whittle's Index policy, i.e.  $\lim_{T \rightarrow \infty} C_T^N(\mathbf{x})/N = C^{WIP,N}/N$ , with the one obtained by applying the Max-Weight policy  $MWP$ ,  $C^{MWP,N}/N$ . The policy  $MWP$  schedules, at each time  $t$ , the  $M$  weighted longest queues (equivalently the  $M$  highest  $a_k q_i^k(t)$ ). We also compare the performance of these two policies with the optimal cost per user obtained by using the optimal solution of the relaxed problem, i.e.  $C^{RP,N}/N$ . The results are plotted in Figures (4.4.a) and (4.4.b) respectively for the initial states  $\mathbf{x}$  and  $\mathbf{y}$  (defined above).

One can see that for large  $N$ , regardless of the initial state, the cost incurred by adopting the Whittle's Index policy tends to the optimal cost of the relaxed problem, which proves that it asymptotically converges to the optimal solution of the original problem. One can also remark that the optimal cost of the relaxed problem per user is constant and does not depend on  $N$  (see section 4.5). Lastly, we remark that the solution given by  $MWP$  is suboptimal and lacks behind our proposed scheduling scheme.

## Fairness among users

In order to improve the fairness among the users in the network, one can use the developed Whittle index policy up to some modifications. To that extent, we introduce in this section a new policy  $\Theta$  which works as follows: at each time slot  $t$ , we schedule the users with the highest  $W_k(q_i^k(t))\overline{D}_k(q_i^k(t))$ , where  $q_i^k(t)$  is the queue state of user  $i$  in class  $k$ ,  $W_k$  is the Whittle index of state  $q_i^k(t)$  when the transmission rate is  $R_k$  and  $\overline{D}_k(q_i^k(t)) = \frac{\sum_{u=1}^t a_k q_i^k(u)}{t}$ . To evaluate numerically the performance of this policy, we consider the following two costs  $C_1^{\pi,N}$  and  $C_2^{\pi,N}$  incurred respectively by users of class 1 and users of class 2 under policy  $\pi$ , specifically  $C_1^{\pi,N} = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^{\gamma_1^N} a_1 q_i^1(t) \mid \mathbf{x}, \pi \right]$  and  $C_2^{\pi,N} = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{i=1}^{\gamma_2^N} a_2 q_i^2(t) \mid \mathbf{x}, \pi \right]$ . We plot these quantities over  $N$  when  $\pi = WIP$  and when  $\pi = \Theta$  with respect to  $N$  in figure 4.5. We conclude that the new policy gives a better performance in terms of fairness, since it reduces the gap between the costs of the two classes of users.

## 4.8.2 Numerical Results for infinite buffer size

For this scenario, we compare the average cost given by our policy  $WIP$  with the one given by the myopic policy or the Max-Weight policy. We plot the results on Figures where we consider two classes of users with their respective transmission rate  $R_1$  and  $R_2$ , and the number of servers is equal to  $N/2$  where  $N$  is the number of users. In Figure (4.6.a), we take  $R_1 = 5$  and  $R_2 = 20$ , while in Figure (4.6.b), the values of  $R_1$  and  $R_2$  are respectively 10 and 45. According to these figures, one can show that our policy  $WIP$  is asymptotically optimal even with  $L = +\infty$ , and performs much better than the myopic policy. This confirms our main motivation behind developing the Whittle index policy when  $L$  is infinite.

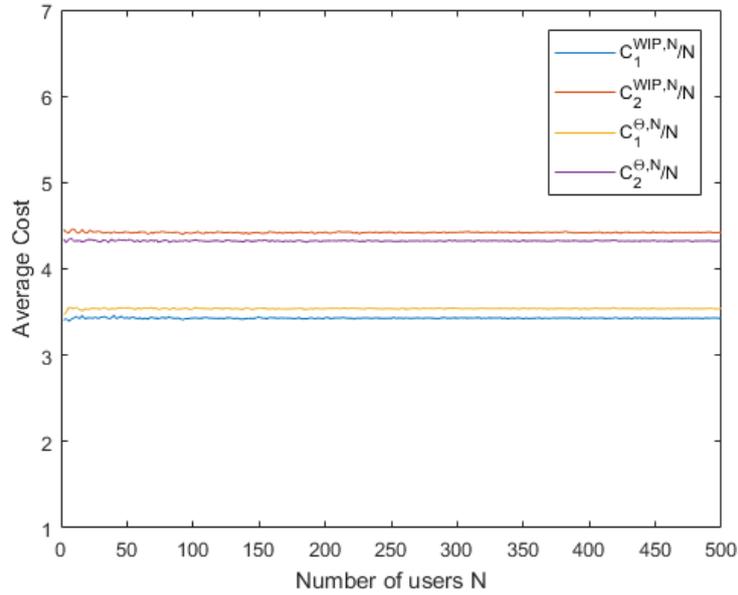


Figure 4.5 – Evaluation of  $C_1^{\pi,N}$  and  $C_2^{\pi,N}$  in function of  $N$  under Policy  $\Theta$  and Whittle's Index policy WIP

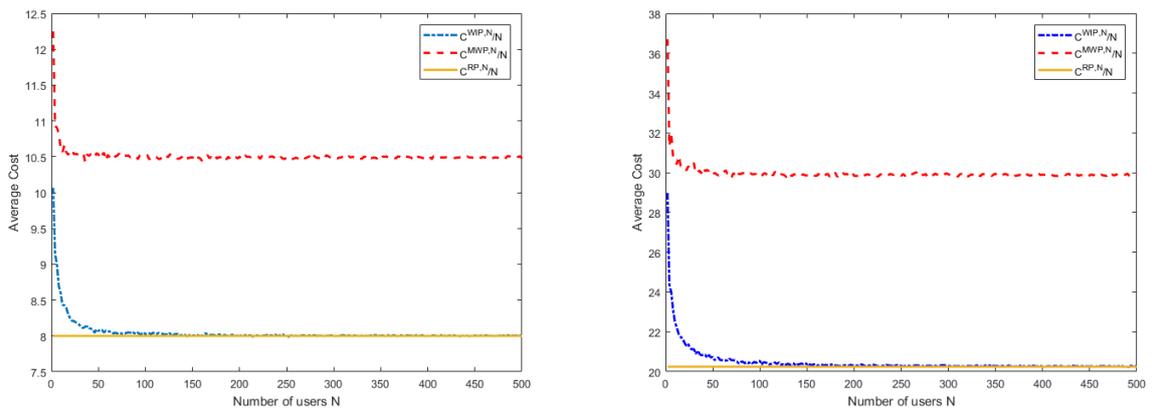


Figure 4.6 – Performance evaluation of Whittle's Index policy

## 4.9 Conclusion

In this chapter, we have studied the problem of users and channels scheduling under bursty traffic arrivals. At each time slot, only  $M$  channels can be allocated to the users knowing that a user can be allocated one channel at most. For both scenarios ( $L < +\infty$  and  $L = +\infty$ ), we have formulated a Lagrangian relaxation of the optimization problem and have provided a characterization of the optimal solution of this relaxed problem. We have then developed a simple Whittle's Index policy to allocate the channels to the users and have proved its asymptotic local and global optimality when the numbers of users and channels are large enough and when the buffer size  $L$  is finite. This result is of interest as the developed Whittle's Index policy has a low complexity and is near optimal for large number of users. We have then provided numerical results that corroborate our claims.

## Chapter 5

# Whittle's index policy for minimizing the Age of Information

In this chapter, we apply Whittle index policy (WIP) for minimizing the age of information (AoI) in the framework of scheduling problems. For that purpose, we start by showcasing our system model in the context of AoI minimization problem and in RBP framework. Then, following the same methodology detailed in Chapter 3, we derive the corresponding Whittle's index policy. Afterwards, we provide the rigorous analytical proof of the local and the global optimality of WIP. Lastly, we give some numerical results that put into perspective the exactitude of our theoretical findings. We emphasize that in this chapter, we show the local optimality of Whittle's index policy for a system containing **several classes**. Whereas, we prove the global optimality for a system containing **two different classes**. We note that to establish the last result, we proceed with a novel and original approach completely different from that used in the previous chapter and in the literature. Essentially, this new method doesn't require that assumption 4.1 be taken into account.

### 5.1 System Model

#### 5.1.1 Network description

We consider  $N$  users that send status updates to a monitor. Due to the limited amount of channels, only  $M < N$  users can transmit simultaneously. Let  $\alpha = \frac{M}{N} \in ]0, 1[$  be the portion of the  $N$  users that can transmit. Time is considered to be discrete. We recall that in our system model, neither the users nor the channels are equipped with buffers, but instead, the channel is considered to be unreliable, which means that: at time slot  $t$ , if user  $i$  is scheduled, then the channel is good, i.e., the transmitted packet is successfully received by the monitor at time  $t + 1$  with a probability  $p_i$ . Otherwise, the packet fails to reach the destination. Moreover, we suppose that users are divided into  $K$  classes such that the probability of successful transmission for users in class  $k$  is  $p_k$ . Each class  $k$  has  $\gamma_k N$  users and, consequently, the following holds:  $\sum_{k=1}^K \gamma_k = 1$ . Our goal is to let the monitor to have the freshest information carried by the users. To that end, we need first to introduce the AoI metric adapted to our system model.

### 5.1.2 Adopted age metric

At each time instant  $t \geq 0$ , we let  $g_i^k(t)$  be the time-stamp of the freshest packet by user  $i$  of class  $k$  that has been delivered to the monitor. Subsequently, the age of information of this user is defined as:

$$S_i^k(t) = \min(t - g_i^k(t), L) \quad (5.1)$$

where  $L$  is in  $\mathbb{N}^* \cup \{+\infty\}$ . To that extent, we study in the sequel the problem of channels allocation with two different metrics depending on  $L$ :

1.  $L$  is finite
2.  $L$  is infinite

To do so, we adapt the methodology given in Chapter 3 for this particular model and similarly to the chapter 4, we derive the theoretical results for both metrics in parallel.

### 5.1.3 Problem formulation

We let the age vector at time  $t$  be  $\mathbf{S}(t) = (S_1^1(t), \dots, S_{\gamma_K^N}^K(t))$  where  $S_i^k(t)$  is the age at the monitor of user  $i$  of class  $k$  at time slot  $t$ . A scheduling policy  $\pi$  is defined as a sequence of actions  $\pi = (\mathbf{a}^\pi(0), \mathbf{a}^\pi(1), \dots)$  where  $\mathbf{a}^\pi(t) = (a_1^{1,\pi}(t), a_2^{1,\pi}(t), \dots, a_{\gamma_K^N}^{K,\pi}(t))$  is a binary vector such that  $a_i^{k,\pi}(t) = 1$  if user  $i$  of class  $k$  is scheduled at time  $t$ . We let  $c_i^{k,\pi}(t) \in \{0, 1\}$  be an i.i.d. Bernoulli random variable that indicates if the transmitted packet by the scheduled user is successfully received (value  $c_i^{k,\pi}(t) = 1$ ) or not (value  $c_i^{k,\pi}(t) = 0$ ). Hence, by definition of  $c_i^{k,\pi}(t)$ , we have that  $\Pr(c_i^{k,\pi}(t) = 1) = p_k$  and  $\Pr(c_i^{k,\pi}(t) = 0) = 1 - p_k$ . Therefore, the evolution of the age of user  $i$  of class  $k$  under policy  $\pi$  can be summarized in the following:

$$S_i^k(t+1) = \begin{cases} 1 & \text{if } a_i^{k,\pi}(t) = 1, c_i^{k,\pi}(t) = 1 \\ \min(S_i^k(t) + 1, L) & \text{otherwise} \end{cases} \quad (5.2)$$

Denoting by  $\Pi$ , the set of all *causal* scheduling policies, our scheduling problem can be formulated as follows:

$$\begin{aligned} & \underset{\pi \in \Pi}{\text{minimize}} && \limsup_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}^\pi \left( \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} S_i^{k,\pi}(t) \mid \mathbf{S}(0) \right) \\ & \text{subject to} && \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} a_i^{k,\pi}(t) \leq \alpha N \quad t = 1, 2, \dots \end{aligned} \quad (5.3)$$

According to Chapter 3, the problem in (5.3) belongs to the family of RMAB problems, then, the optimal solution is out of reach. For this reason, we derive the Whittle's index policy due to its low complexity and good performance. To that end, we follow the same steps explained in Chapter 3. However, in order to avoid redundancy, we skip the part about the Lagrangian Relaxation approach, and we will be limited to give only the important results in the form of Propositions and Theorems.

1. The Bellman equation that corresponds to the dual relaxed problem for a given user after decomposition is:

$$\theta + V(S) = \min_{a \in \{0,1\}} \left\{ S + Wa + \sum_{S' \in \{1, \dots, L\}} \Pr(S \rightarrow S' | a) V(S') \right\} \quad (5.4)$$

where  $\Pr(S \rightarrow S')$  is the transition probability from the age state  $S$  to  $S'$ ,  $\theta$  is the optimal value of the problem and  $V(S)$  is the differential cost-to-go function.

2. Since we are not able to find the explicit solution of the Bellman equation for a given  $W$ , we focus only on studying the structure of the optimal scheduling policy. By doing so, the following results can be obtained. First of all, we recall the definition of an increasing threshold policy.

**Definition 5.1.** *An increasing threshold policy is a policy  $\pi_n \in \Pi$  such that when the age of information is **larger or equal**<sup>1</sup> to  $n$ , the user is scheduled. Otherwise, it is not scheduled.*

**Theorem 5.1.** *The optimal solution of the problem in (5.4) is an increasing threshold policy.*

*Proof.* The proof can be found in Appendix C.1. □

3. In order to establish the indexability of the problem and find the Whittle's index expressions, we tackle in more depth the behavior of the MDP that characterize the evolution of the age when a threshold policy is adopted. In fact, under a threshold policy  $n$ , the MDP can be modeled through a Discrete Time Markov Chain (**DTMC**) as seen in Fig. 5.1 and in Fig. 5.2.

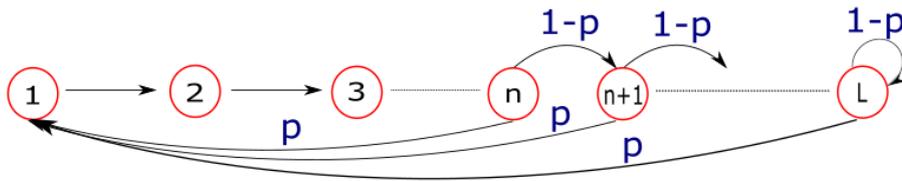


Figure 5.1 – The states transitions under a threshold policy  $n$  when  $L < +\infty$

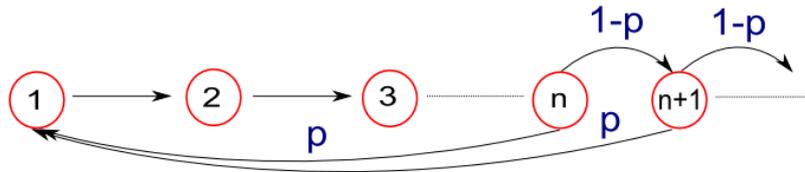


Figure 5.2 – The states transitions under a threshold policy  $n$  when  $L = +\infty$

<sup>1</sup>Unlike the definition of threshold policy given in Chapter 4, here we suppose that under threshold policy  $\pi_n$ , even at state  $n$ , the action prescribed is the active action

To that extent, we derive in the next proposition, the stationary distribution of the DTMC in question for both cases of  $L$ .

**Proposition 5.1.** *For any given threshold  $n \in \{1, \dots, L\}$ , the DTMC is irreducible and admits  $u^n(i)$  for  $i = 1, \dots, L$  as its stationary distribution:*

- *Finite  $L$ :*

(a)  $n \leq L$

$$u^n(i) = \begin{cases} \frac{p}{np+1-p} & \text{if } 1 \leq i \leq n \\ (1-p)^{i-n} \frac{p}{np+1-p} & \text{if } i \geq n \\ \frac{(1-p)^{L-n}}{np+1-p} & \text{if } i = L \end{cases} \quad (5.5)$$

(b)  $n \geq L + 1$

$$u^n(i) = \begin{cases} 0 & \text{if } 1 \leq i \leq L - 1 \\ 1 & \text{if } n = L \end{cases} \quad (5.6)$$

- *Infinite  $L$ :*

$$u^n(i) = \begin{cases} \frac{p}{np+1-p} & \text{if } 1 \leq i \leq n \\ (1-p)^{i-n} \frac{p}{np+1-p} & \text{if } i \geq n \end{cases} \quad (5.7)$$

*Proof.* The expressions results from solving the full balance equation satisfied by  $u^n(\cdot)$  while bearing mind that  $\sum_{i=1}^L u^n(i) = 1$ .  $\square$

Based on the stationary distribution under a threshold policy  $n$ , we can give the steady-state form of the one-dimensional dual relaxed problem that corresponds to the Bellman equation (5.4) as follows:

$$\min_{n \in [1, L]} \left[ \sum_{i=1}^L i u^n(i) + W \sum_{i=n}^L u^n(i) \right] \quad (5.8)$$

where  $W$  is the Lagrangian parameter,  $\sum_{i=1}^L i u^n(i) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}^{\pi_n} \left[ \sum_{t=0}^{T-1} s^{\pi_n}(t) \right]$  refers to the long run expected average age under threshold policy  $\pi_n$ , and  $\sum_{i=n}^L u^n(i) = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}^{\pi_n} \left[ \sum_{t=0}^{T-1} a^{\pi_n}(t) \right]$  indicates the long run expected average time when the active action is prescribed to the user of interest under threshold policy  $\pi_n$ .

As it was mentioned in Chapter 3, this closed form of the one-dimensional dual relaxed problem turns out to be crucial to derive the Whittle index values as we will see in the sequel.

By leveraging the above results, we can now proceed with finding a closed-form of the average age as well as the average active time under a given threshold policy  $\pi_n$  for both cases of  $L$ .

**Theorem 5.2.** *For any given threshold  $n$ , the explicit expressions of  $\sum_{i=1}^L i u^n(i)$  and  $\sum_{i=n}^L u^n(i)$  are:*

- *Finite  $L$ :*

(a)  $n \in [0, L]$ :

$$\sum_{i=1}^L iu^n(i) = \frac{[(n-1)^2 + (n-1)]p^2 + 2p(n-1)}{2p((n-1)p+1)} + \frac{2[1 - (1-p)^{L-n+1}]}{2p((n-1)p+1)} \quad (5.9)$$

$$\sum_{i=n}^L u^n(i) = \frac{1}{np+1-p} \quad (5.10)$$

(b)  $n \geq L+1$ :

$$\sum_{i=1}^L iu^n(i) = L \quad (5.11)$$

$$\sum_{i=n}^L u^n(i) = 1 \quad (5.12)$$

• *Infinite L:*

$$\sum_{i=1}^L iu^n(i) = \frac{[(n-1)^2 + (n-1)]p^2 + 2p(n-1)}{2p((n-1)p+1)} + \frac{2}{2p((n-1)p+1)} \quad (5.13)$$

$$\sum_{i=n}^L u^n(i) = \frac{1}{np+1-p} \quad (5.14)$$

4. Based on the above results, we establish the indexability property of the problem for all users, which ensures the existence of the Whittle's indices and allows us to establish our index policy as it was explained in Chapter 3.

**Proposition 5.2.** *For both cases of  $L$ , and for each user belonging to any class  $k = 1, \dots, K$ , the one-dimensional problem is indexable.*

*Proof.* The proof can be found in Appendix C.2. □

As the indexability property has been established in the above proposition, we can now affirm the existence of the Whittle's index. To that extent, we provide the expressions of Whittle's indices in the following theorem.

**Theorem 5.3.** *The Whittle's index of state  $i \in [1, L]$  in class  $k$  is defined as:*

• *Finite L:*

$$W_i^k = \frac{i(i-1)p_k}{2} + i - i(1-p_k)^{L-i} \quad (5.15)$$

• *Infinite L:*

$$W_i^k = \frac{i(i-1)p_k}{2} + i \quad (5.16)$$

*Proof.* The proof can be found in Appendix C.3.  $\square$

5. In light of the last result, we summarize in the following the Whittle's index scheduling policy for the original problem (5.3).

---

**Algorithm 3** Whittle's index scheduling policy

---

- 1: At each time slot  $t$ , calculate the Whittle's index of all users using Theorem 5.3.
  - 2: Schedule the  $M$  users having the highest Whittle's index values at time  $t$ .
- 

Although the above scheduling policy is easy to implement, it remains sub-optimal. Subsequently, characterizing its performance compared to the optimal policy is important.

## 5.2 Local optimality: Finite $L$

In this section, we study the local optimality properties of the Whittle's index policy when  $L$  is finite. To that end, we reintroduce the state space  $\{Z_i^{k,N}(t), 1 \leq i \leq \gamma_k N, 1 \leq k \leq K\}$  defined in Chapter 3, Section 3.4 over which the local optimality will be established. Analogously to the method used in the previous chapter with regard to the optimality of WIP, we proceed with the fluid approximation technique to establish the desired result. Accordingly, the deterministic vector  $\mathbf{z}(t)$  evolves as:

$$\mathbf{z}(t+1) - \mathbf{z}(t)|_{\mathbf{z}(t)=\mathbf{z}} = \mathbb{E}[\mathbf{Z}^N(t+1) - \mathbf{Z}^N(t)|\mathbf{Z}^N(t) = \mathbf{z}] \quad (5.17)$$

In order to obtain the linear relation between  $\mathbf{z}(t+1)$  and  $\mathbf{z}(t)$  when the Whittle's index policy is implemented, we should first determine the transition probability from a given state of age to another state for a given  $\mathbf{z}(t) = \mathbf{z}$ . To that end, we reintroduce the probability that a user of class  $k$  in state  $i$  is selected among the proportion  $z_i^k$  for transmission:

$$p_i^k(\mathbf{z}) = \min\{z_i^k, \max(0, \alpha - \sum_{W_j^h > W_i^k} z_j^h)\}/z_i^k, \quad (5.18)$$

Next, as it was done in Chapter 4, Section 4.6, we denote by  $q_{i,j}^{k,0}$  and  $q_{i,j}^{k,1}$  the probability of transition from state  $i$  to state  $j$  in a class  $k$  if the user is left to idle or is scheduled for transmission respectively. The expressions of  $q_{i,j}^{k,0}$  and  $q_{i,j}^{k,1}$  can be easily derived from the age dynamics detailed in (5.2). Specifically:

$$q_{i,j}^{k,0} = \begin{cases} 1 & \text{if } j = \min(i+1, L) \\ 0 & \text{otherwise} \end{cases}$$

$$q_{i,j}^{k,1} = \begin{cases} p_k & \text{if } j = 1 \\ 1 - p_k & \text{if } j = \min(i+1, L) \\ 0 & \text{otherwise} \end{cases} \quad (5.19)$$

Therefore, given  $\mathbf{z}$ , the probability of transition between state  $i$  and  $j$  in class  $k$ ,  $q_{i,j}^k(\mathbf{z})$  which can be also seen as the output stream from the state  $i$  to the state  $j$  given  $\mathbf{z}$  for a user at state  $i$ , is as follows:

$$q_{i,j}^k(\mathbf{z}) = p_i^k(\mathbf{z})q_{i,j}^{k,1} + (1 - p_i^k(\mathbf{z}))q_{i,j}^{k,0}. \quad (5.20)$$

Given that, we can summarize the evolution of  $\mathbf{z}(t)$  as follows:

$$z_i^k(t+1) - z_i^k(t) = \sum_{j \neq i} q_{j,i}^k(\mathbf{z}(t))z_j^k(t) - \sum_{i \neq j} q_{i,j}^k(\mathbf{z}(t))z_i^k(t) \quad (5.21)$$

This can be rewritten in the following manner:

$$\mathbf{z}(t+1) = Q(\mathbf{z}(t))\mathbf{z}(t) + \mathbf{z}(t) \quad (5.22)$$

We recall that to establish the asymptotic optimality of Whittle's index policy, according to 3.4, we compare its performance to the average age of the relaxed problem. Precisely, we show that  $C_T^N(\mathbf{x})$  converges to  $C^{RP,N}$  when  $N$  and  $T$  grows where  $C^{RP,N}$  and  $C_T^N(\mathbf{x})$  are the optimal total average age of the relaxed problem and the total expected average age over time duration  $0 \leq t \leq T$  starting from the initial proportion vector  $\mathbf{z}(0) = \mathbf{x}$  when the Whittle's index policy is adopted respectively. For that, we need to have the expression of  $C^{RP,N}$ . To proceed in that direction, we rewrite the Proposition 4.9 adapted for our specific case.

**Proposition 5.3.** *There exists a class  $m$ , state  $p$ , and a randomization parameter  $\theta^*$  such that the optimal solution of the dual problem when the langrangian parameter  $W = W_p^m = W^*$  is characterized by:*

- For  $k \neq m$ , the optimal threshold is  $l_k(W^*) = \arg \max_i \{W_i^k | W_i^k \leq W_p^m\} + 1$
- For  $k = m$ , the optimal solution is randomized policy between two threshold policies  $l_m(W^*) = \arg \max_i \{W_i^m | W_i^m \leq W_p^m\} + 1$  and  $l_m(W^*) - 1 = \arg \max_i \{W_i^m | W_i^m \leq W_p^m\}$ , where the factor of randomization  $\theta^*$  is the probability of adopting the policy  $l_m(W^*)$  and  $1 - \theta^*$ , the probability of adopting the policy  $l_m(W^*) - 1$ .
- The relaxed constraint is satisfied with equality, i.e.

$$\alpha = \sum_{k \neq m} \sum_{i=l_k(W^*)}^L \gamma_k u_k^{l_k(W^*)}(i) + \sum_{i=l_m(W^*)}^L \gamma_m u_m^*(i) + (1 - \theta^*) \gamma_m u_m^{l_m(W^*)-1}(l_m(W^*) - 1)$$

$$\text{Where } u_m^* = \theta^* u_m^{l_m(W^*)} + (1 - \theta^*) u_m^{l_m(W^*)-1}.$$

**Remark 5.1.** *In the above proposition, one can notice that the thresholds are differed by one compared with that of Proposition 4.9. This comes from the fact that the definition of a threshold policy is not the same for both chapters 4 and 5.*

According to Proposition 5.3, the optimal solution of the relaxed problem is a threshold policy characterized by  $W^*$  and  $\theta^*$  defined in the above proposition, Accordingly,

$$C^{RP,N} = \sum_{\substack{k=1 \\ k \neq m}}^K \gamma_k N \sum_{i=1}^L u_k^{l_k(W^*)}(i)i + \gamma_m N \theta^* \sum_{i=1}^L u_m^{l_m(W^*)}(i)i + \gamma_m N (1 - \theta^*) \sum_{i=1}^L u_m^{l_m(W^*)-1}(i)i \quad (5.23)$$

For ease of notations we denote  $l_k(W^*)$  by  $l_k^*$ . For the local optimality, we restrict our analysis to a specific set in  $\mathcal{Z}$  which is a subset of  $\mathcal{J}_{W^*}$  defined in Section 4.6, Chapter 4. In fact, in this aforementioned set,  $Q(\mathbf{z}(t))$  doesn't depend on  $\mathbf{z}(t)$ . This will be confirmed in the sequel by establishing the link between  $\mathbf{z}(t+1)$  and  $\mathbf{z}(t)$  within this set.

1)  $k \neq m$ : In this case, we can replace  $z_{l_k^*-1}^k(t)$  by  $\gamma_k - \sum_{j=1, j \neq l_k^*-1}^L z_j^k(t)$ . Therefore, there is no need to track the evolution of  $z_{l_k^*-1}^k(t)$  with time as it can be deduced from the evolution of  $z_j^k(t)$  for  $j \neq l_k^* - 1$ . Accordingly, we let  $\tilde{\mathbf{z}}^k(t) = [z_1^k(t), \dots, z_{l_k^*-2}^k(t), z_{l_k^*}^k(t), \dots, z_L^k(t)]$ . Next, by replacing  $q_{j,i}^k(z(t))$  with its value in (5.21), we can obtain the following expression of  $z_i^k(t+1)$  in function of  $\mathbf{z}^k(t)$ :

$$\begin{aligned} & \sum_{j=l_k^*}^L p_k z_j^k(t) && \text{if } i = 1 \\ & z_{i-1}^k(t) && \text{if } 1 < i < l_k^* - 1 \\ & - \sum_{j=1}^{l_k^*-2} z_j^k(t) - \sum_{j=l_k^*}^L z_j^k(t) + \gamma_k && \text{if } i = l_k^* \\ & (1 - p_k) z_{i-1}^k(t) && \text{if } l_k^* < i < L \\ & (1 - p_k) z_{L-1}^k(t) + (1 - p_k) z_L^k(t) && \text{if } i = L \end{aligned} \quad (5.24)$$

2)  $k = m$ : To tackle this case, we first replace  $z_{l_m^*}^m(t)$  by  $\gamma_m - \sum_{j=1, j \neq l_m^*}^L z_j^m(t)$ . Similarly to the first case, there is no need to track the evolution of  $z_{l_m^*}^m(t)$  with time. Accordingly, we let  $\tilde{\mathbf{z}}^m = [z_1^m, \dots, z_{l_m^*-1}^m(t), z_{l_m^*+1}^m(t), \dots, z_L^m(t)]$ . The difference between this case and the above one is that there exists a given set of users in class  $m$  whose Whittle's index equals to  $W^*$ . That implies that, as  $\mathbf{z}(t) \in \mathcal{J}_{W^*}$ , there is a group of users among this aforementioned set which will be scheduled, while the remaining group will be in the idle mode. Thus, unlike the previous case, the relation between  $z_1^m(t+1)$  and  $\mathbf{z}^m(t+1)$  will not be easy to express. Nevertheless, we recall that the portion of scheduled users is always equal to  $\alpha$ . Hence, the portion of scheduled users of class  $m$  can be always written as the difference  $\alpha - \sum_{k \neq m} \sum_{j=l_k^*}^L z_j^k(t)$ . With that in mind, we can write the evolution of  $z_i^m(t+1)$  for any  $i \neq l_m^*$  in function of  $\mathbf{z}(t)$  as follows:

$$\begin{aligned} & (\alpha - \sum_{k \neq m} \sum_{j=l_k^*}^L z_j^k(t)) p_m && \text{if } i = 1 \\ & z_{i-1}^m(t) && \text{if } 1 < i < l_m^* \\ & - \sum_{j=1}^{l_m^*-1} z_j^m(t) - \sum_{j=l_m^*+1}^L (1 - p_m) z_j^k(t) + \gamma_m - p_m (\alpha - \sum_{k \neq m} \sum_{j=l_k^*}^L z_j^k(t)) && \text{if } i = l_m^* + 1 \\ & (1 - p_m) z_{i-1}^m(t) && \text{if } l_m^* + 1 < i < L \\ & (1 - p_m) z_{L-1}^m(t) + (1 - p_m) z_L^m(t) && \text{if } i = L \end{aligned} \quad (5.25)$$

Based on this, we can conclude that when  $\mathbf{z}(t) \in \mathcal{J}_{W^*}$ ,  $\tilde{\mathbf{z}}(t+1)$  and  $\tilde{\mathbf{z}}(t)$  are related through the simpler linear equation:

$$\tilde{\mathbf{z}}(t+1) = Q\tilde{\mathbf{z}}(t) + \mathbf{c}, \quad (5.26)$$

Where the expression of  $Q$  has the following form:

$$Q = \begin{bmatrix} Q_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & Q_2 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & & \\ A_1 & A_2 & \cdots & Q_m & \cdots & A_{K-1} & A_K \\ \vdots & & & \ddots & & \vdots & \\ 0 & 0 & \cdots & \cdots & \cdots & Q_{K-1} & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & Q_K \end{bmatrix}. \quad (5.27)$$

Let us define  $\mathbf{z}^* \in \mathcal{Z}$  as the system state vector that results from adopting the optimal policy of the relaxed problem. By exploiting the results of Propositions 5.1 and 5.3, and using the definition of  $\mathbf{z}^*$ , one can find the exact expression of  $\mathbf{z}^*$ . Specifically:

$$\begin{aligned} z_i^{*,k} &= \gamma_k u_k^{l^*}(i) \\ z_i^{*,m} &= \gamma_m [\theta u_m^{l^*}(i) + (1 - \theta) u_m^{l^*-1}(i)] \end{aligned} \quad (5.28)$$

To prove that  $C_T^N(\mathbf{x})$  converges to  $C^{RP,N}$  when  $T$  and  $N$  scale, we should first demonstrate that  $\mathbf{z}(t)$ , when the Whittle's index policy is adopted, converges to  $\mathbf{z}^*$  within a given subset of  $\mathcal{J}_{W^*}$ . In fact,  $\mathbf{z}^*$  is not simply a system state that belongs to the set  $\mathcal{J}_{W^*}$ , but beyond that, it is the fixed point of the fluid approximation equation (5.17). Bearing that in mind, we have this following always holds when  $\mathbf{z}(t) \in \mathcal{J}_{W^*}$ :

$$\tilde{\mathbf{z}}(t) - \tilde{\mathbf{z}}^* = \mathbf{Q}e \quad (5.29)$$

where  $\mathbf{e} = \tilde{\mathbf{z}}(0) - \tilde{\mathbf{z}}^*$ . As consequence, as it was depicted in Section 4.6 Chapter 4, we must set out that all eigenvalues of  $\mathbf{Q}$  are strictly less than one to establish our desired result. The proof of this statement is elaborated in C.4.

**Proposition 5.4.** *For all eigenvalue  $\lambda$  of  $\mathbf{Q}$ ,  $|\lambda| < 1$*

*Proof.* See appendix C.4. □

Accordingly, there exists  $\sigma > 0$  such that, if  $\mathbf{z}(0) \in \Omega_\sigma(\mathbf{z}^*) \subseteq \mathcal{J}_{W^*}$ , we have  $\mathbf{z}(t) \in \mathcal{J}_{W^*}$  and  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$ . Leveraging these above results, we prove that  $\mathbf{Z}^N(t)$  evolves closely to  $\mathbf{z}^*$  when  $N$  is large.

**Proposition 5.5.** *There exists a neighborhood  $\Omega_\delta(\mathbf{z}^*)$  such that, for any  $T, \mu > 0$ , if  $\mathbf{Z}^N(0) = \mathbf{x} \in \Omega_\delta(\mathbf{z}^*)$ , there exists a constant  $C_1$  independent of  $N$  and  $\mathbf{x}$  such that:*

$$\Pr_{\mathbf{x}}(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu) \leq \frac{C_1}{N} \quad (5.30)$$

where  $\Pr_{\mathbf{x}}$  denotes the probability conditioned on the initial state  $\mathbf{Z}^N(0) = \mathbf{x}$ .

Providing that  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$  within a neighborhood  $\Omega_\delta(\mathbf{z}^*)$ , we get the following corollary based on the previous proposition.

**Corollary 5.1.** *There exists a neighborhood  $\Omega_\delta(\mathbf{z}^*)$  such that, for any  $\mu > 0$ , if  $\mathbf{Z}^N(0) = \mathbf{x} \in \Omega_\delta(\mathbf{z}^*)$ , then there exists a time  $T_0$  such that for any time instant  $T > T_0$ , there exists a constant  $C_f$  independent of  $N$  and  $\mathbf{x}$  such that:*

$$\Pr_{\mathbf{x}}(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu) \leq \frac{C_f}{N} \quad (5.31)$$

*Proof.* The proof is identical to that of Lemma 4.6. □

Based on this corollary, we are ready now to establish the local optimality of the Whittle's index policy.

**Proposition 5.6.** *If the initial state  $\mathbf{x}$  is in the set  $\Omega_\delta(\mathbf{z}^*)$ , then*

$$\lim_{T \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} = \frac{C^{RP,N}}{N} \quad (5.32)$$

*Proof.* The proof is identical to that of Proposition 4.12. Accordingly, it will be skipped for sake of shortness.  $\square$

### 5.3 Global Optimality: Infinite $L$

In this section, we will show that the Whittle Index Policy is asymptotically globally optimal when  $L$  is infinite. Specifically, we show that the Whittle index Policy is optimal for a large number of users  $N$  and a large number of channels  $M$ , the ratio  $\alpha = \frac{M}{N}$  being constant and starting from any initial state. For that, we will compare the average cost obtained by the Whittle Index Policy with the optimal cost of the original problem.

Providing that the number of states of the age is infinite, we denote by  $m^k(t)$ , the highest state at time  $t$  in class  $k$ . Without loss of generality, we have that  $\sum_{i=0}^{m^k(t)} Z_i^{k,N}(t) = \gamma_k$  for each class  $k$ .

We establish the global optimality for two different classes of users where  $p_1$  and  $p_2$  are the successful transmission probabilities of the class 1 and 2 respectively ( $p_1 > p_2$ ). As it has been proven in 5.2, we prove likewise that  $\mathbf{Z}^N(t)$ , evolving under the Whittle index policy, converges in probability to  $\mathbf{z}^*$  when  $N$  and  $t$  are very large. To that extent, we start by showing that the fluid approximation of  $\mathbf{Z}^N(t)$  denoted by  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$ . To that end, for a sake of clarity, we give in the sequel an insight into the theoretical analysis developed to demonstrate this desired result:

Recalling that the relation that links  $\mathbf{z}(t+1)$  to  $\mathbf{z}(t)$  is not linear in general as it was reported in Section 5.2, then establishing the convergence of  $\mathbf{z}(t)$  requires a new method involving terms of  $\alpha_1(t)$  and  $\alpha_2(t)$  where these two users' proportions refer to the scheduled users' proportion at time  $t$  of the class 1 and 2 respectively ( $\alpha_1(t) + \alpha_2(t) = \alpha$ ). To that extent, we give in the following the outlines of the proof:

1. For a large enough time  $t$ , based on Lemma 5.1, we show that there exists  $T_t$  such that we can find a partial relation between each element of the vector  $\mathbf{z}(t+T_t)$  and terms of the sequence  $\{\alpha_k(t')\}_{k=1,2}^{t' \leq t+T_t}$ . More precisely, we prove that for  $T_t$ , we can express each proportion of users that are not scheduled at time  $t+T_t$  in function of one term of  $\{\alpha_k(t')\}_{k=1,2}^{t' \leq t+T_t}$ . This allows us to obtain  $1 - \alpha$  as a linear combination between the terms of  $\{\alpha_k(t')\}_{k=1,2}^{t' \leq t+T_t}$  at time  $t+T_t$ .
2. We define in Definition 5.3,  $T_{\max}$  that satisfies these two following properties proven in Propositions 5.8 and 5.9 using Lemma 5.2:
  - The Whittle index alternates between the two classes from state 1 to  $T_{\max} + 1$  under a given assumption on  $\alpha$ .
  - The instantaneous thresholds  $l_1(\cdot)$  and  $l_2(\cdot)$  are bounded by  $T_{\max}$  at time  $t+T_t$ .

3. Based on the above result, we derive the relation between the instantaneous thresholds at time  $t + T_t$  in Proposition 5.10.
4. Taking as an initial time  $t + T_t = T_0$ , we show by induction in Proposition 5.11, that for all  $T \geq T_0$ , the instantaneous thresholds are less than  $T_{\max}$  and that all proportions containing the non scheduled users can be expressed in function of terms of the sequence  $\{\alpha_k(t')\}_{\substack{k=1,2 \\ t' \leq T}}$ .
5. We define for each class  $k$  a vector  $\mathbf{A}_k(T)$  composed by  $\alpha_k(T)$  (the scheduled users' proportion at time  $T$ ) plus the finite subset of the sequence  $\{\alpha_k(t')\}_{t' \leq T}$  such that for all proportion of users in class  $k$  at a given state at time  $T$  that is not scheduled can be expressed by one element belonging to this subset. Then, we provide the relation between the elements of the vectors  $\mathbf{A}_k(T)$  and  $\mathbf{A}_k(T + 1)$  in Propositions 5.12 and 5.13
6. We conclude the convergence of the highest and the smallest element of  $\mathbf{A}_k(T)$  when  $T$  grows in Theorem 5.4.
7. We demonstrate by contradiction that the highest and the smallest element of  $\mathbf{A}_k(T)$  must converge to the same limit in Proposition 5.14. This last result implies that  $\alpha_k(t)$  converges when  $t$  scales.
8. In light of that fact, we prove that  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$  in Proposition 5.15.
9. Using Kurth theorem, we show in Proposition 4.11 that  $\mathbf{Z}^N(t)$  converges to  $\mathbf{z}^*$  in probability.
10. Finally we establish in Proposition 5.17, the convergence of  $C_T^N(x)$  to  $C^{RP,N}$  as  $N$  and  $T$  grow starting from any initial point  $\mathbf{x}$ .

**Remark 5.2.** *We highly emphasize that the proportion  $\alpha_k(t)$  and  $1 - \alpha$  refer to the scheduled users' proportion at time  $t$  in class  $k$  and the non scheduled users' proportion either for class 1 or 2 respectively. Meanwhile, for any other proportion  $A$ , it refers only to the number of users in this proportion over the total users' number of the system whatever the different states of users that contains. Having said that,  $A = B$  means that they are equal in terms of proportion, while they can contain users in different states.*

We remind that the fluid limit technique consists of analyzing the evolution of the expectation of  $\mathbf{Z}^N(t)$  under the Whittle Index Policy. For ease of reading, we rewrite the equation (5.17):

$$\mathbf{z}(t + 1) - \mathbf{z}(t)|_{\mathbf{z}(t)=\mathbf{z}} = \mathbb{E} [\mathbf{Z}^N(t + 1) - \mathbf{Z}^N(t) | \mathbf{Z}^N(t) = \mathbf{z}] \quad (5.33)$$

This above equation reveals to us that we have a sequence  $\mathbf{z}(t)$  defined by recurrence for a fixed initial state  $\mathbf{z}(0)$  that we should study its behavior when  $t$  is very large. Hence, we end up with a function  $\mathbf{z}(t)$  that depends on two variables,  $t$  and the initial value  $\mathbf{z}(0)$ . To that extent, our aim is to prove that  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$  regardless of the initial state  $\mathbf{z}(0)$  where  $\mathbf{z}^*$  denotes the users' proportion corresponding to the optimal policy of the relaxed problem when  $L$  is infinite. We let  $\mathbf{z}(t) = (\mathbf{z}^1(t), \dots, \mathbf{z}^K(t))$  with

$\mathbf{z}^k(t) = (z_1^k(t), \dots, z_{m^k(t)}^k(t))$  where  $z_i^k(t)$  is the expected proportion of users at state  $i$  in class  $k$  at time  $t$  with respect to the equation (5.33). Accordingly we have that  $\sum_{i=0}^{m^k(t)} z_i^k(t) = \gamma_k$  for each class  $k$ .

One can notice that  $\mathbf{z}^*$  is a particular vector with respect to the equation (5.33).

**Proposition 5.7.**  $\mathbf{z}^*$  is the **unique** fixed point of the fluid approximation equation. In other words,  $\mathbf{z}(t) = \mathbf{z}(t+1)$ , if and only if  $\mathbf{z}(t) = \mathbf{z}^*$ .

*Proof.* The proof follows the same methodology of the paper [32, Lemma 9] □

According to this proposition, it is sufficient to show that  $\mathbf{z}(t)$  converges starting from any initial state  $\mathbf{z}(0)$ , as the only eventual finite limit of  $\mathbf{z}(t)$  when  $t$  tends to  $+\infty$  is the fixed point of the equation (5.33),  $\mathbf{z}^*$ .

In the following, we prove that the fluid approximation vector of  $\mathbf{Z}^N(t)$ ,  $\mathbf{z}(t)$  under the Whittle Index Policy converges starting from any initial state. We prove this result for 2 different classes of users where  $p_1$  and  $p_2$  are the successful transmission probabilities of the class 1 and 2 respectively ( $p_1 > p_2$ ), given a sufficient condition on  $\alpha$ . Throughout this section, we denote by  $w_1(n)$  and  $w_2(n)$ , the Whittle's index, whose expression is given in Proposition 5.3, of state  $n$  in class 1 and class 2 respectively. We need to prove that  $z_i^k(t)$  converges for each state  $i$  in class  $k$ .

Now, focusing on the Whittle index policy, we can see it as an instantaneous threshold policy for each class, where the thresholds vary over time  $t$ . Moreover, under the Whittle index policy, the proportion of users that are scheduled at each time slot  $t$  is fixed and equals to  $\alpha$  since the number of scheduled users at each time slot  $t$  is  $\alpha N$ . This proportion  $\alpha$  contains the users with the highest Whittle index values. In that respect, we define  $\alpha_1(t)$  and  $\alpha_2(t)$  the proportion of users in class 1 and class 2 respectively at time  $t$  with the highest Whittle index values such that  $\alpha_1(t) + \alpha_2(t) = \alpha$ . The remaining proportion of users which are not scheduled at each time slot  $t$ , which is equal to  $1 - \alpha$ , contains the users with the smallest Whittle index values. Now, regarding this proportion, we give its decomposition into proportions of users at different states in different classes. Denoting by  $l_1(t)$  and  $l_2(t)$  at time  $t$  the instantaneous threshold integers under Whittle index policy, then there exists two real values between 0 and 1,  $\beta(t)$  and  $\gamma(t)$ , with  $\gamma(t) = 1$  and  $0 < \beta(t) \leq 1$ , or  $0 < \gamma(t) \leq 1$  and  $\beta(t) = 1$ , such that:

$$\sum_{i=1}^{l_1(t)-1} z_i^1(t) + \sum_{i=1}^{l_2(t)-1} z_i^2(t) + \beta(t)z_{l_1(t)}^1(t) + \gamma(t)z_{l_2(t)}^2(t) = 1 - \alpha \quad (5.34)$$

and  $\{z_i^1(t)\}_{1 \leq i \leq l_1(t)} \cup \{z_i^2(t)\}_{1 \leq i \leq l_2(t)}$  is exactly the set  $\{z_i^k : w_k(i) \leq \max(w_1(l_1(t)), w_2(l_2(t)))\}$ . For the case where  $L$  is finite, we have assumed that  $\mathbf{z}(0)$  is within a neighborhood of  $\mathbf{z}^*$ . This assumption has allowed us to find an easy linear relation between  $\mathbf{z}(t)$  and  $\mathbf{z}(t+1)$  ( $\mathbf{z}(t+1) = \mathbf{Q}\mathbf{z}(t) + \mathbf{c}$ ), and then deduce the convergence of  $\mathbf{z}(\cdot)$  by establishing that the spectral value of  $\mathbf{Q}$  is strictly less than one as it was depicted in Proposition 5.4. Whereas for this case, when  $L$  is infinite, since we aim to prove the convergence of  $\mathbf{z}(\cdot)$  from any initial state, the general relation between  $\mathbf{z}(t+1)$  and  $\mathbf{z}(t)$  is as follows:

$$\mathbf{z}(t+1) = \mathbf{Q}(\mathbf{z}(t))\mathbf{z}(t) + \mathbf{c}(t) \quad (5.35)$$

This equation is not linear which makes studying the evolution of  $\mathbf{z}(\cdot)$  a hard task. Moreover, as the number of state is infinite, then the dimensions of  $\mathbf{z}(t)$  varies per time. Therefore, the matrix  $\mathbf{Q}(\mathbf{z}(t))$  is not square. Hence we can not apply the same method as in Section 5.2 since the spectral values are not defined for a non square matrix. For these reasons, we proceed differently than 5.2. Our method consists in fact on expressing each proportion  $z_i^k(t)$  that belongs to a non scheduled users' proportion at time  $t$  in function of a term of  $\alpha_k(\cdot)$  at a given time less than  $t$ . By this way, we will obtain a part of the vector  $\mathbf{z}(t)$  in function of  $\{\alpha_k(t')\}_{\substack{t' \leq t, \\ k \in \{0,1\}}}$ , and the sum of the other part equal to  $\alpha$ . Then, we show that  $\alpha_k(\cdot)$  converges for  $k = 1, 2$ . We will see later that it is sufficient to show that  $\alpha_k(\cdot)$  converges in order to conclude for the convergence of  $\mathbf{z}(\cdot)$ . To find the partial relation between  $\mathbf{z}(t)$  and  $\{\alpha_k(t')\}_{\substack{t' \leq t, \\ k \in \{0,1\}}}$ , we prove the following lemma.

**Lemma 5.1.** *Knowing  $\mathbf{z}^k(t)$ ,  $\alpha_k(t)$  and  $l_k(t)$ , we have that:*

*For  $i = 1$ :*

$$z_1^k(t+1) = p_k \alpha_k(t).$$

*For  $1 \leq i < l_k(t)$ :*

$$z_{i+1}^k(t+1) = z_i^k(t).$$

*Proof.* See appendix C.5. □

According to Lemma 5.1, after scheduling under the Whittle's Index Policy, we get at time  $t+1$ , a proportion of  $p_1 \alpha_1(t)$  of users at state 1 in class 1 and  $p_2 \alpha_2(t)$  of users at state 1 in class 2 respectively (i.e.  $z_1^1(t+1) = p_1 \alpha_1(t)$  and  $z_1^2(t+1) = p_2 \alpha_2(t)$ ).

According to the same lemma, at time  $t+2$ , a proportion of  $p_1 \alpha_1(t)$  and  $p_2 \alpha_2(t)$  of users will go to state 2 in class 1 and class 2 respectively and  $p_1 \alpha_1(t+1)$ ,  $p_2 \alpha_2(t+1)$  of users will move to state 1 in class 1 and class 2 respectively (i.e.  $z_1^1(t+2) = p_1 \alpha_1(t+1)$ ,  $z_1^2(t+2) = p_2 \alpha_2(t+1)$ ,  $z_2^1(t+2) = p_1 \alpha_1(t)$  and  $z_2^2(t+2) = p_2 \alpha_2(t)$ ).

At time  $t+3$ , a proportion of  $p_1 \alpha_1(t)$  and  $p_2 \alpha_2(t)$  of users will go to state 3 in class 1 and class 2 respectively,  $p_1 \alpha_1(t+1)$ ,  $p_2 \alpha_2(t+1)$  of users will move to state 2 in class 1 and class 2 respectively,  $p_1 \alpha_1(t+2)$  and  $p_2 \alpha_2(t+2)$  of users will move to state 1 in class 1 and class 2 respectively, (i.e.  $z_1^1(t+3) = p_1 \alpha_1(t+2)$ ,  $z_1^2(t+3) = p_2 \alpha_2(t+2)$ ,  $z_2^1(t+3) = p_1 \alpha_1(t+1)$  and  $z_2^2(t+3) = p_2 \alpha_2(t+1)$ ,  $z_3^1(t+3) = p_1 \alpha_1(t)$ ,  $z_3^2(t+3) = p_2 \alpha_2(t)$ ).

Thereby, at time  $t+t_0$  where the instantaneous threshold  $l_k(t+t_0) \geq t_0$ , we get a set of proportions

$\{p_1 \alpha_1(t), p_2 \alpha_2(t), \dots, p_1 \alpha_1(t+t_0-1), p_2 \alpha_2(t+t_0-1)\}$  that belong to the proportion  $1 - \alpha$  of users with the lowest Whittle index values, such that  $z_1^1(t+t_0) = p_1 \alpha_1(t+t_0-1)$ ,  $z_1^2(t+t_0) = p_2 \alpha_2(t+t_0-1)$ ,  $\dots$ ,  $z_{t_0}^1(t+t_0) = p_1 \alpha_1(t)$  and  $z_{t_0}^2(t+t_0) = p_2 \alpha_2(t)$ . Hence, we obtain a  $z_i^k(t+1)$  which is well expressed in function of terms of  $\alpha_k(\cdot)$  ( $k = 1, 2$ ) for  $i \in [1, t_0]$ ,  $k = 1, 2$ .

**Remark 5.3.** *Considering Whittle index policy framework, the order of the different users' proportions with respect to their Whittle index values must be taking into account throughout this analysis. In fact, as we have already mentioned, we need to give the expression of the non scheduled users' proportions in function of the terms of  $\alpha_k(\cdot)$  for  $k = 1, 2$ , which can not be done only if we consider the order of the Whittle index values. To that extent, since the set of the non scheduled users' proportions, according to the Whittle's index policy, is exactly the set of users' proportions with the lowest Whittle*

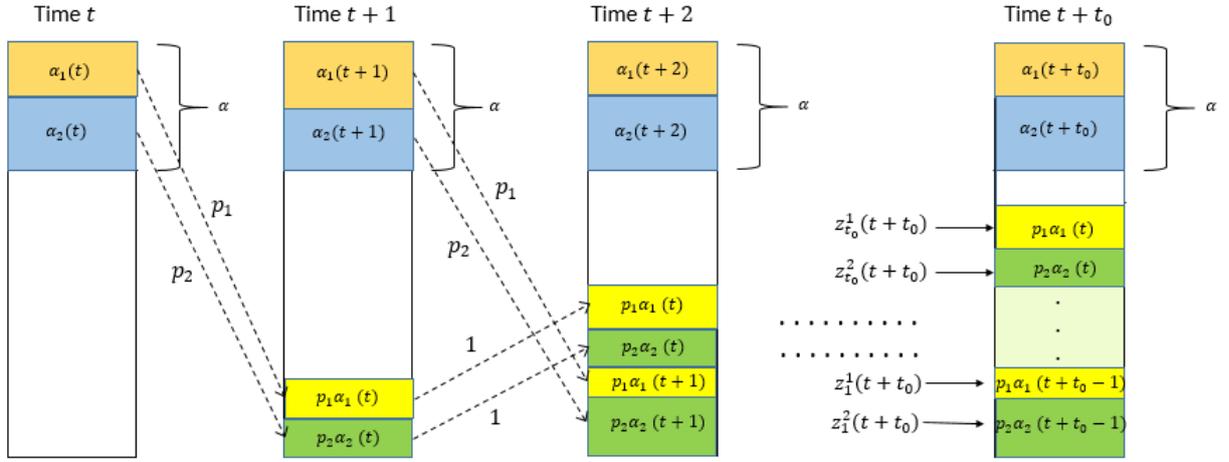


Figure 5.3 – Evolution of  $z_i^k(\cdot)$  for different states  $i$  in function of  $\alpha_1(t)$  and  $\alpha_2(t)$  under the Whittle Index Policy (the green and the yellow colors refer to class 1 and 2 respectively)

index values among all the different users' proportions of the system, then the form at time  $t$  of this specific set will be  $\{z_i^k(t) : w_k(i) \leq w_m(n)\}$  for a given  $m$  and  $n$  that vary with  $t$ .

Based on this remark above, we need to find at time  $t + t_0$ , a set of the form  $\{z_i^k(t + t_0) : w_k(i) \leq w_m(n)\}$  for a given class  $m$  and state  $n$ , such all the elements of this set are well expressed in function of  $\alpha_k(\cdot)$ . We show in the sequel that the highest Whittle index of this set could be  $w_2(t_0)$ .

Indeed, given that the Whittle index function is increasing with  $n$  where  $n$  refers to a given age of information state, then for any state in class 2 with Whittle index less than  $w_2(t_0)$ , belongs to  $[1, t_0]$ . Moreover, considering the state  $q$  in class 1 such that  $w_1(q) \leq w_2(t_0) \leq w_1(t_0)$  ( $p_1 > p_2$ ), then  $w_1(q) \leq w_1(t_0)$ , which means that  $q \in [1, t_0]$ . Hence, for any element in  $\{z_i^k(t + t_0) : w_k(i) \leq w_2(t_0)\}$ , can be expressed in function of terms of  $\alpha_k(\cdot)$  ( $k = 1, 2$ ). Accordingly,  $\{z_i^k(t + t_0) : w_k(i) \leq w_2(t_0)\}$  equals to the set  $\{p_2\alpha_2(t), \dots, p_2\alpha_2(t + t_0 - 1), p_1\alpha_1(t + t_0 - l(t + t_0)), \dots, p_1\alpha_1(t + t_0 - 1)\}$ , where  $l(t + t_0)$  is the greatest state  $q$  in class 1 such that  $w_1(q) \leq w_2(t_0)$ . We note that  $l(t + t_0) \leq t_0$  because  $w_2(l(t + t_0)) \leq w_1(l(t + t_0)) \leq w_2(t_0)$ .

Therefore, in that regards, for a fixed  $t$ , we associate for each  $t_0$  the corresponding sum  $\sum_{j=1}^{l(t+t_0)} z_j^1(t + t_0) + \sum_{j=1}^{t_0} z_j^2(t + t_0) = \sum_{j=1}^{l(t+t_0)} p_1\alpha_1(t + t_0 - j) + \sum_{j=1}^{t_0} p_2\alpha_2(t + t_0 - j)$ . To that extent, we define in the following the time  $t_0$  when this aforementioned sum exceeds  $1 - \alpha$ .

**Definition 5.2.** Starting at time  $t$ , we define  $T_t$  such that  $t + T_t$  is the first time that verifies:

$$\sum_{j=1}^{l(t+T_t)} p_1\alpha_1(t + T_t - j) + \sum_{i=1}^{T_t} p_2\alpha_2(t + T_t - j) \geq 1 - \alpha \quad (5.36)$$

In other words, the first time when  $\sum_{j=1}^{l(t+t_0)} p_1\alpha_1(t + t_0 - j) + \sum_{i=1}^{t_0} p_2\alpha_2(t + t_0 - j)$  exceeds  $1 - \alpha$  is  $t + t_0 = t + T_t$ .

Then, at time  $t + T_t$ , there exists  $l'_1(t + T_t) \leq l(t + T_t)$ ,  $l'_2(t + T_t) \leq T_t$ , such that the set  $\{z_i^1(t + T_t)\}_{1 \leq i \leq l'_1(t + T_t)} \cup \{z_i^2(t + T_t)\}_{1 \leq i \leq l'_2(t + T_t)}$  is exactly the set  $\{z_i^k(t + T_t) : w_k(i) \leq \max(w_1(l'_1(t + T_t)), w_2(l'_2(t + T_t)))\}^2$ , and  $\gamma(t + T_t) = 1$  and  $0 < \beta(t + T_t) \leq 1$ , or  $0 < \gamma(t + T_t) \leq 1$  and  $\beta(t + T_t) = 1$  such that:

$$\begin{aligned} & \sum_{j=1}^{l'_1(t+T_t)-1} p_1 \alpha_1(t + T_t - j) + \sum_{j=1}^{l'_2(t+T_t)-1} p_2 \alpha_2(t + T_t - j) \\ & + \beta(t + T_t) p_1 \alpha_1(t + T_t - l'_1(t + T_t)) + \gamma(t + T_t) p_2 \alpha_2(t + T_t - l'_2(t + T_t)) = 1 - \alpha, \end{aligned} \quad (5.37)$$

with  $l'_1(t + T_t)$  and  $l'_2(t + T_t)$  being the instantaneous thresholds in class 1 and 2 respectively at time  $t + T_t$ .  $\alpha_1(t + T_t)$  and  $\alpha_2(t + T_t)$  are the users' proportions with the highest Whittle index values, and their sum is equal to  $\alpha$ . Without loss of generality, we let  $l'_k(t + T_t) = l_k(t + T_t)$ .

As we can see, at time  $t + T_t$ , all the expressions of the users' proportions that belong to the  $1 - \alpha$  of users with the smallest Whittle index values, are in function of  $\alpha_1(t)$  or  $\alpha_2(t)$  at various time. In fact, at time  $t + T_t$ , we end up with  $z_1^1(t + T_t) = p_1 \alpha_1(t + T_t - 1)$ ,  $z_1^2(t + T_t) = p_2 \alpha_2(t + T_t - 1)$ ,  $\dots$ ,  $z_{l'_1(t+T_t)}^1(t + T_t) = p_1 \alpha_1(t + T_t - l_1(t + T_t))$  and  $z_{l'_2(t+T_t)}^2(t + T_t) = p_2 \alpha_2(t + T_t - l_2(t + T_t))$ , and the rest of the proportions belongs to  $\alpha_1(t + T_t)$  for class 1 and  $\alpha_2(t + T_t)$  for class 2. For this reason, we work only with  $\alpha_1(\cdot)$  and  $\alpha_2(\cdot)$  in order to prove the convergence. As we have mentioned earlier, the proof of the optimality is valid under an assumption on  $\alpha$ . This later relies on the maximum value that can take the instantaneous thresholds  $l_k(t + T_t)$  at time  $t + T_t$  for  $k = 1, 2$ . To that extent, we start by defining and bounding a certain constant  $T_{\max}$ . Then under an assumption on  $\alpha$ , we show that the order of Whittle index alternates between the two classes in the set  $[1, T_{\max} + 1]$  (this will be detailed later). Based on this, we establish that  $T_{\max}$  is an upper bound of  $l_k(t + T_t)$ .

First of all, we give a lemma which will be useful to prove the propositions 5.9, 5.10 and 5.11.

**Lemma 5.2.** *There exists a time  $t_f$  such that for all  $t \geq t_f$ ,  $\alpha_1(t) > 0$ .*

*Proof.* See appendix C.6. □

In this following definition, we define  $T_{\max}$ , and we check later that it coincides with the upper bound of  $l_k(t + T_t)$  for  $k = 1, 2$ .

**Definition 5.3.** *Starting at time  $t$ , we define  $T_{\max}$  as  $T_t$  defined in Definition 5.2, that verifies the following:*

- $\sum_{j=1}^{l(t+T_t)} p_1 \alpha_1(t + T_t - j) + \sum_{j=1}^{T_t} p_2 \alpha_2(t + T_t - j) \geq 1 - \alpha$
- $\alpha_1(t + i) = 0$  for all  $i \in [0, T_{\max} - 1]$

In the next lemma, we determine the upper and the lower bound of  $T_{\max}$ .

**Lemma 5.3.**  *$T_{\max}$  doesn't depend on  $t$  and satisfies:  $\frac{1-\alpha}{p_2\alpha} \leq T_{\max} \leq \frac{1-\alpha}{p_2\alpha} + 1$ .*

<sup>2</sup>According to Remark 5.3, the form of this set means that it contains the users' proportions with the lowest Whittle index values among all users' proportions of the system

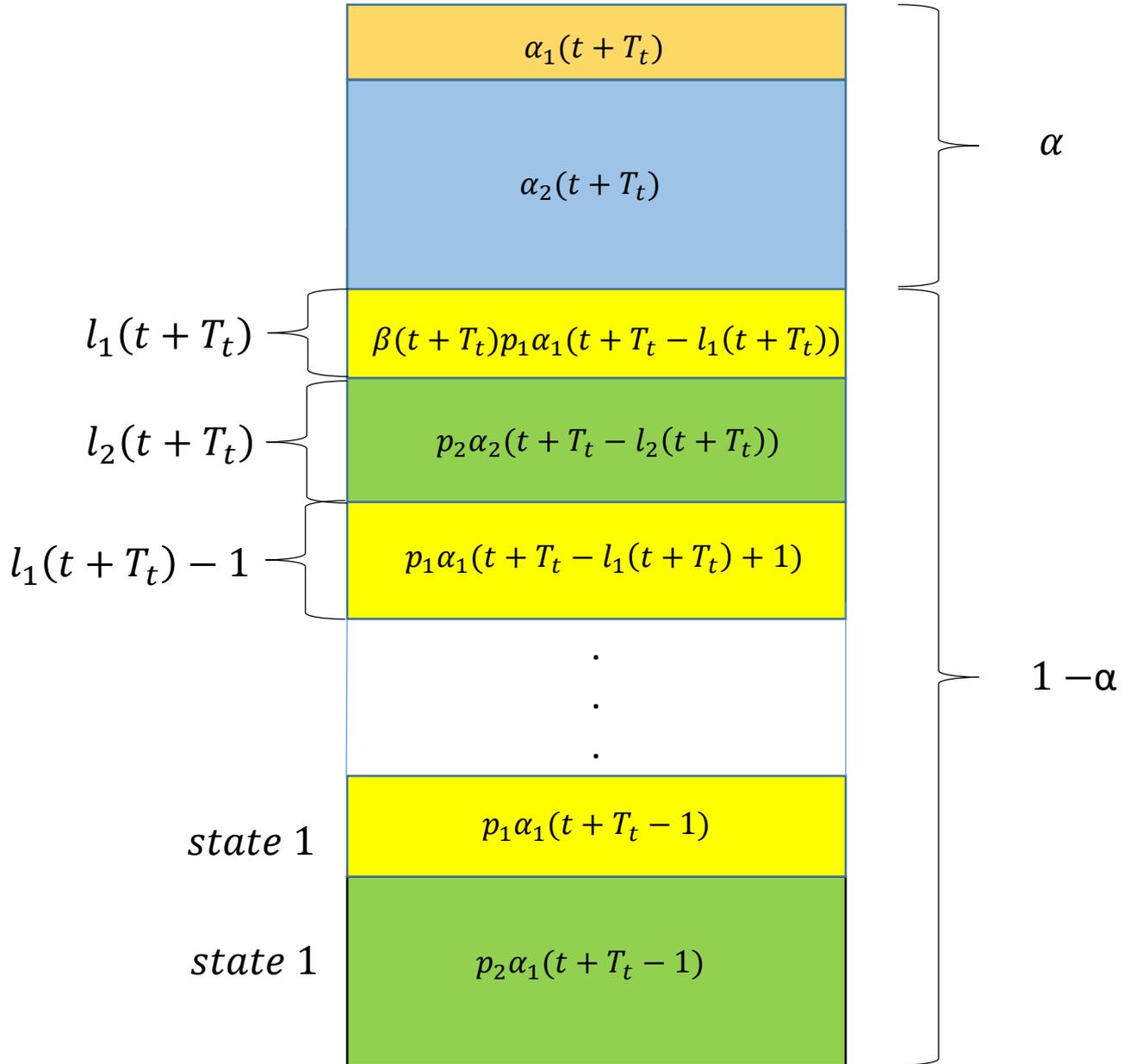


Figure 5.4 – The proportions of users at different states at time  $t + T_t$  when  $\gamma(t + T_t) = 1$  and  $0 < \beta(t + T_t) \leq 1$

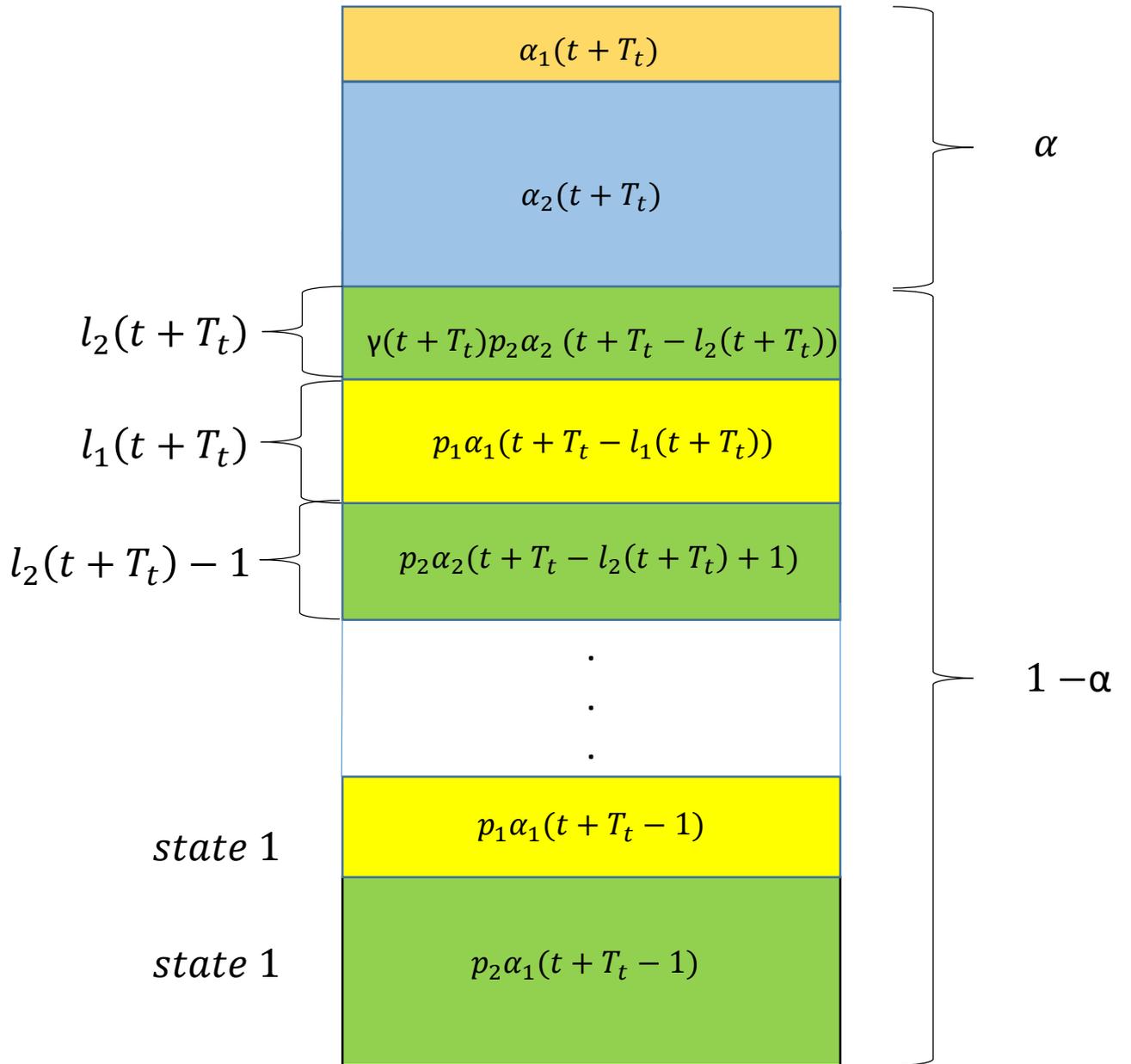


Figure 5.5 – The proportions of users at different states at time  $t + T_t$  when  $\beta(t + T_t) = 1$  and  $0 < \gamma(t + T_t) \leq 1$



*Proof.* See appendix C.7. □

We say that the order of the Whittle index strictly alternates between the two classes in  $[1, n]$  or from state 1 to  $n$ , if we have  $w_2(1) < w_1(1) < w_2(2) < w_1(2) < w_2(3) < w_1(3) < \dots < w_2(n) < w_1(n)$ . To that extent, the proof of  $\alpha_k(\cdot)$  convergence is feasible when the alternation condition is satisfied from 1 to  $l_k(t + T_t) + 1$  for all  $t$ . We note that this condition will be relevant in the proof of the proposition 5.13. To that end, we start by introducing the assumption on  $\alpha$ . Then, we demonstrate effectively that under this assumption the condition of alternation is satisfied from 1 to  $l_k(t + T_t) + 1$ .

**Assumption 5.1.** Denoting  $\frac{1}{p_1 - p_2} \left( \frac{p_1 + p_2}{2} + \sqrt{2(p_1 - p_2) + \frac{(p_1 + p_2)^2}{4}} \right)$  by  $D$ . Then, the users' proportion scheduled at each time  $\alpha$  satisfies:

$$\alpha > \frac{1}{1 + (D - 2)p_2} \quad (5.38)$$

If  $T_{\max}$  is the highest value that  $l_k(t + T_t)$  can take, (this will be shown later in proposition 5.9), then it is sufficient to prove that the hypothesis of the Whittle index alternation is satisfied from 1 to  $T_{\max} + 1$ . This will be shown in the next proposition.

**Proposition 5.8.** Under Assumption (5.1), the order of the Whittle index alternates between the two classes from state 1 to state to  $T_{\max} + 1$ .

*Proof.* See appendix C.8. □

Now we prove that the instantaneous thresholds of the two classes can not exceed  $T_{\max}$ .

**Proposition 5.9.** Denoting by  $l_{\max}$  the highest instantaneous threshold in the sense that  $\forall t \geq t_f, \max(l_1(t + T_t), l_2(t + T_t)) \leq l_{\max}$ , then  $T_{\max} = l_{\max}$ .

*Proof.* See appendix C.9 □

According to the last proposition,  $T_{\max}$  is truly the upper bound of  $l_k(t + T_t)$  for all  $t$  and  $k = 1, 2$ . As consequence, the order of the Whittle index alternates between the two classes in the set  $[1, l_k(t + T_t) + 1]$ . The next goal is to find a relation between  $l_1(t + T_t)$  and  $l_2(t + T_t)$ . To do so, we recall that we have at time  $t$ :

$$\sum_{i=1}^{l_1(t)-1} z_i^1(t) + \sum_{i=1}^{l_2(t)-1} z_i^2(t) + \beta(t)z_{l_1(t)}^1(t) + \gamma(t)z_{l_2(t)}^2(t) = 1 - \alpha \quad (5.39)$$

with  $l_1(t)$  and  $l_2(t)$  being the thresholds in class 1 and 2 respectively at time  $t$ , and  $\beta(t) = 1$  and  $0 < \gamma(t) \leq 1$ , or  $\gamma(t) = 1$  and  $0 < \beta(t) \leq 1$ . Thereby, the first step consists of establishing the relationship between  $l_1(t)$  and  $l_2(t)$  when  $\max(l_1(t), l_2(t)) \leq T_{\max}$  depending on two different cases that we will explain thereafter in order to give a generalized expression of the aforementioned equation (5.39) where the index of the class is not specified in the expressions of the thresholds  $l_1(t)$  and  $l_2(t)$ .

**Remark 5.4.** It is worth mentioning that, as we have defined  $l_{\max}$  in Proposition 5.9, it refers to the highest value that can be attained by the thresholds of the class 1 or 2 at time  $t + T_t$  for  $t > t_f$  where  $t_f$  is a given in Lemma 5.2. Whereas, at any time  $t > t_f$ ,  $\max(l_1(t), l_2(t)) \leq l_{\max}$  might not be true since we don't have necessary a given  $t'$  such that  $t' + T_{t'} = t$  for any  $t > t_f$ .

**Proposition 5.10.** At any time  $t > t_f$ , if  $\max(l_1(t), l_2(t)) \leq T_{\max} = l_{\max}$ , then there exists  $l(t) \leq l_{\max}$  and,  $\beta(t) = 0$  and  $0 < \gamma(t) \leq 1$ , or  $0 < \beta(t) \leq 1$  and  $\gamma(t) = 1$  such that:

$$\sum_{i=1}^{l(t)-1} z_i^1(t) + \sum_{i=1}^{l(t)-1} z_i^2(t) + \beta(t)z_{l(t)}^1(t) + \gamma(t)z_{l(t)}^2(t) = 1 - \alpha \quad (5.40)$$

*Proof.* See appendix C.11. □

Starting at time  $t \geq t_f$ , we have that at time  $t + T_t$ , the thresholds  $l_1(t + T_t)$  and  $l_2(t + T_t)$  are less than  $l_{\max}$ . Hence, according to Proposition (5.10), there exists  $l(t + T_t)$  such that:

$$\begin{aligned} \sum_{j=1}^{l(t+T_t)-1} p_1\alpha_1(t + T_t - j) + \sum_{j=1}^{l(t+T_t)-1} p_2\alpha_2(t + T_t - j) \\ + \beta(t + T_t)p_1\alpha_1(t + T_t - l(t + T_t)) + \gamma(t + T_t)p_2\alpha_2(t + T_t - l(t + T_t)) = 1 - \alpha \end{aligned} \quad (5.41)$$

where  $\beta(t + T_t) = 0$  and  $0 \leq \gamma(t + T_t) < 1$ , or  $0 \leq \beta(t + T_t) < 1$  and  $\gamma(t + T_t) = 1$ .

Denoting  $t + T_t$  by  $T_0$ , we obtain:

$$\sum_{j=1}^{l(T_0)-1} p_1\alpha_1(T_0 - j) + \sum_{j=1}^{l(T_0)-1} p_2\alpha_2(T_0 - j) + \beta(T_0)p_1\alpha_1(T_0 - l(T_0)) + \gamma(T_0)p_2\alpha_2(T_0 - l(T_0)) = 1 - \alpha \quad (5.42)$$

where  $\beta(T_0) = 0$  and  $0 < \gamma(T_0) \leq 1$ , or  $0 < \beta(T_0) \leq 1$  and  $\gamma(T_0) = 1$ .

Now, we prove by induction that this latter expression is valid for all  $T \geq T_0$ , and that  $l(T)$ , the instantaneous threshold at time  $T$ , is less than  $l_{\max}$ .

**Proposition 5.11.** For all  $T \geq T_0$ , there exists  $l(T) \leq l_{\max}$ ,  $\beta(T)$  and  $\gamma(T)$ , such that:

$$\sum_{j=1}^{l(T)-1} p_1\alpha_1(T - j) + \sum_{j=1}^{l(T)-1} p_2\alpha_2(T - j) + \beta(T)p_1\alpha_1(T - l(T)) + \gamma(T)p_2\alpha_2(T - l(T)) = 1 - \alpha \quad (5.43)$$

where  $\beta(T) = 0$  and  $0 < \gamma(T) \leq 1$ , or  $0 < \beta(T) \leq 1$  and  $\gamma(T) = 1$ .

*Proof.* See appendix C.12. □

According to the latter proposition, we can now define at each time  $T \geq T_0$ , for each class  $k$ , the vector  $\mathbf{A}_k(T) = (\alpha_k(T), \alpha_k(T - 1), \dots, \alpha_k(T - l(T)))$ , such that, there exists  $\beta(T)$

and  $\gamma(T)$ :

$$\sum_{j=1}^{l(T)-1} p_1 \alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2 \alpha_2(T-j) + \beta(T) p_1 \alpha_1(T-l(T)) + \gamma(T) p_2 \alpha_2(T-l(T)) = 1 - \alpha \quad (5.44)$$

where  $\beta(T) = 0$  and  $0 < \gamma(T) \leq 1$ , or  $0 < \beta(T) \leq 1$  and  $\gamma(T) = 1$ . We note that as we have explained previously, the relation between  $\mathbf{A}_k(T)$  and  $\mathbf{z}^k(T)$  is:  $p_k \alpha_k(T-1) = z_1^k(T)$ ,  $p_k \alpha_k(T-2) = z_2^k(T)$ ,  $\dots$ ,  $p_k \alpha_k(T-l(T)) = z_{l(T)}^k(T)$ .

**Remark 5.5.** We emphasize that in the following analysis,  $T$  is always considered greater than  $T_0$ .

We prove in the sequel that  $\max \mathbf{A}_k(T)$  is decreasing and  $\min \mathbf{A}_k(T)$  is increasing (with the max and min referring to the element of the vector with the greatest value, and the smallest value respectively). After that, we conclude the convergence of  $\max \mathbf{A}_k(T)$  and  $\min \mathbf{A}_k(T)$  when  $T$  tends to  $+\infty$ . Then, we prove that they must converge to the same real number. In order to prove that  $\max \mathbf{A}_k(T)$  is decreasing and  $\min \mathbf{A}_k(T)$  is increasing, we first demonstrate this following proposition.

**Proposition 5.12.** All the elements of the vector  $\mathbf{A}_k(T+1)$  belong to the elements of the vector  $\mathbf{A}_k(T)$  except  $\alpha_k(T+1)$ .

*Proof.* See appendix C.13 □

With the intention of proving the monotony of  $\max \mathbf{A}_1(T)$  and  $\min \mathbf{A}_1(T)$ , we still need to prove that the value of  $\alpha_1(T+1)$  must be less than  $\max \mathbf{A}_1(T)$  and greater than  $\min \mathbf{A}_k(T)$ . For that, we introduce the following proposition.

Before doing that, we note that, as  $\alpha_1(t) + \alpha_2(t) = \alpha$  at each time slot  $t$ , then it is sufficient for us to prove that  $\alpha_1(\cdot)$  is converging. To that extent, we study only the vector function  $\mathbf{A}_1(T)$  in order to prove the convergence.

**Proposition 5.13.** Under assumption 5.1, for a given vector  $\mathbf{A}_1(T) = (\alpha_1(T), \alpha_1(T-1), \dots, \alpha_1(T-l(T))) (T \geq T_0)$ , we have four possible cases of inequalities:

$$\alpha_1(T) \leq \alpha_1(T+1) \leq \alpha_1(T-l(T))$$

$$\alpha_1(T-l(T)) \leq \alpha_1(T+1) \leq \alpha_1(T)$$

$$\alpha_1(T-l(T)+1) \leq \alpha_1(T+1) \leq \alpha_1(T)$$

$$\alpha_1(T) \leq \alpha_1(T+1) \leq \alpha_1(T-l(T)+1)$$

Moreover: If  $\alpha_1(T) \leq \alpha_1(T+1) \leq \alpha_1(T-l(T))$ , then:

$$\alpha_1(T+1) - \alpha_1(T) \leq p_1(\alpha_1(T-l(T)) - \alpha_1(T))$$

If  $\alpha_1(T-l(T)) \leq \alpha_1(T+1) \leq \alpha_1(T)$ , then:

$$\alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)))$$

If  $\alpha_1(T - l(T) + 1) \leq \alpha_1(T + 1) \leq \alpha_1(T)$ , then:

$$\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T) + 1))$$

If  $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1)$ , then:

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T))$$

*Proof.* See appendix C.14. □

**Theorem 5.4.**  $\min \mathbf{A}_1(T)$  and  $\max \mathbf{A}_1(T)$  converge and we denote their limits respectively by  $l_1$  and  $l_2$ .

*Proof.* According to Proposition 5.12, the elements of the vector  $\mathbf{A}_1(T+1)$  except the first element which is  $\alpha_1(T+1)$  belong to the elements of the vector  $\mathbf{A}_1(T)$ . Hence, the values of these elements (except the first element of  $\mathbf{A}_1(T+1)$ ) is less than  $\max \mathbf{A}_1(T)$  and greater than  $\min \mathbf{A}_1(T)$ . According to the first result of Proposition 5.13, we deduce that  $\alpha_1(T+1)$  is between two values of two elements of the vector  $\mathbf{A}_1(T)$ . Hence, combining the results of Proposition 5.12 and 5.13,  $\max \mathbf{A}_1(T+1) \leq \max \mathbf{A}_1(T)$  and  $\min \mathbf{A}_1(T+1) \geq \min \mathbf{A}_1(T)$ . Then  $\max \mathbf{A}_1(T)$  is decreasing with  $T$  and  $\min \mathbf{A}_1(T)$  is increasing with  $T$ . Given that for all  $T$ ,  $0 \leq \alpha_1(T) \leq \alpha$ , then  $\max \mathbf{A}_1(T)$  and  $\min \mathbf{A}_1(T)$  are bounded by 0 and  $\alpha$ . Therefore, we can conclude that  $\min \mathbf{A}_1(T)$  and  $\max \mathbf{A}_1(T)$  converge and we denote their limits by  $l_1$  and  $l_2$  respectively. Moreover  $\max \mathbf{A}_1(T)$  is lower bounded by  $l_2$  and  $\min \mathbf{A}_1(T)$  is upper bounded by  $l_1$ . □

However, in order to have  $\alpha_1(T)$  converges to a unique point, we need to establish that  $\max \mathbf{A}_1(T)$  and  $\min \mathbf{A}_1(T)$  converge to the same limit. In other words, we need to prove that  $l_1 = l_2$ . For that, we will use the second result of Proposition 5.13. To that extent, we proceed by contradiction, i.e. we suppose that  $l_1 \neq l_2$ . More specifically, given that  $l_1 \leq l_2$  by definition, the two possible cases satisfied by  $l_1$  and  $l_2$  are:  $l_1 < l_2$  or  $l_1 = l_2$ , then to show that  $l_1 = l_2$ , it is sufficient to find a contradiction considering  $l_1 < l_2$ .

In fact, we prove that if  $l_1 < l_2$ , there exists  $T_d$  such that all the elements of  $\mathbf{A}_1(T_d)$  are strictly less than  $l_2$ , that contradicts with the fact that  $\max \mathbf{A}_1(T)$  is lower bounded by  $l_2$ .

As  $\max \mathbf{A}_1(T)$  converges to  $l_2$ , then for a given  $\epsilon > 0$ , there exists a given time slot that we denote by  $T_\epsilon \geq T_0$  such that for all  $T \geq T_\epsilon$ ,  $\max \mathbf{A}_1(T) < l_2 + \epsilon$ . Our proof consists of showing that for a small enough  $\epsilon$ , there exists  $T \geq T_\epsilon$ ,  $\max \mathbf{A}_1(T)$  is strictly less than  $l_2$ . We need first to determine an upper bound of the number of the elements of the vector  $\mathbf{A}_1(T)$  whatever  $T$ . In fact, as we have demonstrated that at each time  $T$ , the instantaneous threshold  $l(T)$  is less than  $l_{\max}$ . Then the number of the elements of  $\mathbf{A}_1(T)$  will not exceed  $l_{\max} + 1$ . In the following proof, we denote  $l_{\max}$  by  $L$ .

**Proposition 5.14.** If  $l_1 < l_2$ , for  $\epsilon \leq (l_2 - l_1) \frac{(1-p_1)^L}{1-(1-p_1)^L}$ , there exist  $T_d \geq T_\epsilon$  such that all the elements of  $\mathbf{A}_1(T_d)$  are strictly less than  $l_2$ .

*Proof.* See appendix C.15. □

Providing that  $l_2$  is a lower bound of  $\max \mathbf{A}_1(T)$  which contradicts with the result of the above proposition. Hence, the supposition of  $l_1 \neq l_2$  is not valid.

Therefore,  $l_1 = l_2$ . Consequently,  $\max \mathbf{A}_1(T)$  and  $\min \mathbf{A}_1(T)$  converge to the same limit denoted  $\alpha_1^*$ . Given that  $\min \mathbf{A}_1(T) \leq \alpha_1(T) \leq \max \mathbf{A}_1(T)$  for all  $T$ , then  $\alpha_1(T)$  also converges to  $\alpha_1^*$ . Similarly,  $\alpha_2(T)$  converges to  $\alpha - \alpha_1^* = \alpha_2^*$ . In the following proposition, we prove that  $\mathbf{z}(t)$  converges.

**Proposition 5.15.** *If  $\alpha_k(t)$  converges to  $\alpha_k^*$ , then for each state  $i$  and class  $k$ ,  $z_i^k(t)$  converges to  $z_i^{k,*}$ .*

*Proof.* See appendix C.16. □

However, we still have to establish that the stochastic vector  $\mathbf{Z}^N(t)$  converges to  $\mathbf{z}^*$  in probability when  $N$  scales. For that, we introduce the following proposition inspired from the discrete-time version of Kurtz Theorem in [59]. Before that, knowing that the norms on the infinite dimension vector space are not equivalents, we work only with a specific norm which will be useful to show the optimality of the Whittle index's policy. Accordingly, we define  $\|\cdot\|$  as follows:

$$\|\mathbf{v}\| = \sum_{i=1}^{+\infty} |v_i^1| i + \sum_{i=1}^{+\infty} |v_i^2| i \quad (5.45)$$

where  $v_i^k$  is the  $i$ -th component in the class  $k$  of the vector  $\mathbf{v}$ . The reason behind chosen a such norm will be revealed in the proof of Proposition 5.17.

**Proposition 5.16.** *For any  $\mu > 0$  and finite time horizon  $T$ , there exists positive constant  $C$  such that*

$$P_{\mathbf{x}}\left(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu\right) \leq \frac{C}{N}$$

where  $P_{\mathbf{x}}$  denotes the probability conditioned on the initial state  $\mathbf{Z}^N(0) = \mathbf{z}(0) = \mathbf{x}$ . Furthermore,  $C$  is independent of  $N$ .

*Proof.* See appendix C.17. □

According to the Proposition above, the system state  $\mathbf{Z}^N(t)$  behaves very close to the fluid approximation model  $\mathbf{z}(t)$  when the number of users  $N$  is large and starting from any initial state. To that extent, in order to establish the optimality of Whittle's index policy, we give first this following lemma which is a consequence of the Proposition 5.16.

**Lemma 5.4.** *For any  $\mu > 0$ , there exists a time  $T_0$  such that for each  $T > T_0$ , there exists a positive constant  $s$  with,*

$$P_{\mathbf{x}}\left(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu\right) \leq \frac{s}{N}$$

*Proof.* See appendix C.18. □

We remind that starting from an initial state  $\mathbf{x}$ , our objective is to compare the total expected average age per user under Whittle index policy which can be expressed as  $\frac{1}{T}\mathbb{E}^{wi}\left[\sum_{t=0}^{T-1}\sum_{k=1}^K\sum_{i=1}^{+\infty}Z_i^{k,N}(t)i\mid\mathbf{Z}^N(0)=\mathbf{x}\right]$  where  $\mathbf{Z}^N(t)$  evolves under Whittle index policy, with the optimal age of the relaxed problem per user whose expression in function of  $\mathbf{z}^*$  is,  $C^{RP}=\frac{C^{RP,N}}{N}=\sum_1^K\sum_{i=1}^{+\infty}z_i^{k,*}i$ , when the number of users  $N$  as well as the time duration  $T$  grow.

According to Lemma 5.4, we are ready now to establish the asymptotic optimality of the Whittle index policy.

**Proposition 5.17.** *Starting from a given initial state  $\mathbf{Z}^N(0)=\mathbf{z}(0)=\mathbf{x}$ , then:*

$$\lim_{T\rightarrow+\infty}\lim_{N\rightarrow\infty}\frac{1}{T}\mathbb{E}^{wi}\left[\sum_{t=0}^{T-1}\sum_{k=1}^K\sum_{i=1}^{+\infty}Z_i^{k,N}(t)i\mid\mathbf{Z}^N(0)=\mathbf{x}\right]=\sum_{k=1}^K\sum_{i=1}^{+\infty}z_i^{k,*}i\quad(5.46)$$

*Proof.* See appendix C.19. □

## 5.4 Numerical Results

### Verification of Assumption 5.1

In this section, we compute the value of the lower bound on  $\alpha$  given in Assumption 5.1. We denote this lower bound by  $B_\alpha$ . For a wide range of parameters  $p_1$  and  $p_2$ , we provide an exhaustive table that represents the lower bound on  $\alpha$  in function of  $p_1$  and  $p_2$ . As can be seen, the lower bound decreases when  $p_1$  and  $p_2$  are close one to the other. Moreover, it grows even smaller when  $p_1$  and  $p_2$  have relatively high values. According to table 5.1, we can notice that in most cases of  $(p_1, p_2)$ , the lower bound  $B_\alpha$  doesn't exceed 0.5. This implies that the interval of  $\alpha$  where the assumption 5.1 is satisfied, is enough wide for different values of  $p_1$  and  $p_2$ .

$p_1$	0.1	0.2	0.3	0.4	0.4	0.5	0.5	0.6	0.7	0.8
$p_2$	0.2	0.4	0.5	0.6	0.8	0.8	1	0.9	0.9	0.9
$B_\alpha$	0.7034	0.6250	0.4711	0.3556	0.5328	0.3612	0.5	0.2893	0.1675	0.1351

Table 5.1 – Evaluation of  $B_\alpha$  under wide range of channel statistics

### Implementation of the Whittle's index policy

In this section, we evaluate the performance of the Whittle's index policy by comparing the per user average age of the Whittle's index policy to the optimal per user average age of the relaxed problem  $C^{RP,N}/N$ . For that purpose, we consider two scenarios, when  $L$  is finite, and precisely equal to 100 and when  $L$  is infinite<sup>3</sup>. We further consider for both scenarios that  $\frac{N}{2}$  is the number of users of the class 1 and class 2 where  $N$  is the total number of users. Accordingly,  $\gamma_1=\gamma_2=\frac{1}{2}$ . The probability of a successful transmission of class 1 and class 2 are respectively  $p_1=0.8$  and  $p_2=0.5$ . At each time slot  $t$ , at most,  $M=\frac{N}{2}$  of users can be scheduled per each time slot, therefore  $\alpha=\frac{M}{N}=\frac{1}{2}$ .

<sup>3</sup>We adopt the Whittle's index policy derived in Theorem 5.3 for both scenarios.

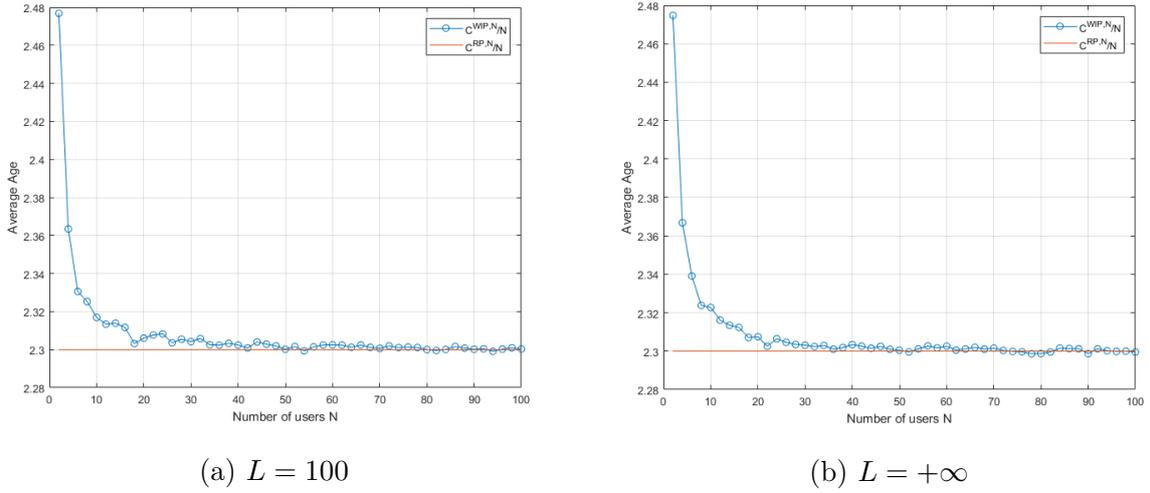


Figure 5.8 – Average age per user under Whittle Index Policy

As seen in Figure 5.8, the performance gap between the Whittle's index policy and the lower-bound  $C^{RP,N}/N$  shrinks as the number of users increases and vanishes for high values of  $N$  for both cases (a) and (b). These results showcase that, besides its optimality in the symmetric case as shown in [43] (i.e., when  $p_k = p$  for  $k = 1, \dots, K$ ), the Whittle's index policy is also optimal for the general asymmetric case in the many-users regime.

## 5.5 Conclusion

In this paper, we have examined the average age minimization problem where only a fraction of the network users can transmit simultaneously over unreliable channels. We have provided analytical results on the local optimality of the Whittle's index policy in the many-users regime for finite  $L$ . While for infinite  $L$ , we presented and derived a novel method based on Cauchy criterion to prove the Whittle's index policy's global optimality. Numerical results were then presented that corroborate our theoretical findings and showcase the optimal performance of the policy when  $N$  grows.



# Chapter 6

## Conclusions and Outlook

### 6.1 Conclusion

In this thesis, we have focused on user scheduling frameworks that minimize the average age of information and average delay. Specifically, we have studied in Chapter 4, a scheduling problem in which one base station allocates the available channels to a subset of users in such a way to minimize the expected total average delay metric incurred in the users' queues. Based on Lagrangian relaxation approach, we derive the Whittle indices for the case where the queue size is finite and the case where the queue size is infinite. We also provide rigorous analysis to establish the asymptotic optimality of Whittle's index policy when the queue size is finite. To that extent, we prove the local asymptotic optimality using fluid approximation techniques, and the global asymptotic optimality under a recurrence assumption. To highlight the performance of Whittle's index policy, we give some numerical results that effectively affirm our theoretical findings, i.e., WIP is asymptotically optimal and in addition, it performs much better than the myopic policy. In Chapter 5, we have examined a resource allocation problem in the context of AoI where one base station allocates the unreliable channels to a subset of users or sensors in such a way to minimize the expected total average age incurred in the sensors. Similarly to the chapter 4, we follow the Lagrangian relaxation approach to derive the Whittle's index policy for two different cases: the first one is when the age value can not exceed a certain upper bound, while the second one is when the age value can take any integer value. Our main contributions reside in showing the optimality of Whittle index policy. Indeed, we prove that WIP is asymptotically locally optimal for a system model composed of several classes when the age is bounded. Whereas, when the age is unbounded, we prove that WIP is asymptotically globally optimal for a system model composed of two classes using an original and novel method. As for numerical results, we have evaluated the performance of WIP by comparing it with the optimal solution of the relaxed problem. Our obtained simulations show that WIP is definitively optimal for a network system with high density, i.e., when the number of users and channels grows.

### 6.2 Future Works

For future works, there exists several scheduling problems with various system settings that can be examined for which we can derive the low-complex and the asymptotically

optimal policy: WIP.

In Chapter 4, we have investigated a multi-class system where the packets arrival follows an uniform distribution and the departure rate is constant over time. The model and analysis of this chapter can be extended by considering:

- General packets arrival distribution.
- Unreliable channels: the transmission rate process is a sequence of an i.i.d (independent and identically distributed) Bernoulli random variables.

Another interesting case to be examined is the Poissonian packet arrival distribution when the buffer size is infinite. The ultimate question arises when dealing with these type of scenarios is whether a closed-form expressions of Whittle indices can be obtained or not. Indeed, deriving WIP is not always feasible for any system model and requires considering assumptions on the parameters of the system. Meanwhile, for a general system model, one can at least look for an algorithm with low complexity that allows to obtain a general expressions of Whittle indices.

Regarding the method applied to prove the global asymptotic optimality of WIP in the chapter 5, one can apply it for different contexts and system models, namely the one studied in the chapter 4. Furthermore, one can extend the proof for several classes instead of two classes and where the assumption 5.1 is violated.

Moreover, in our thesis we consider a linear objective function of delay or age. One can derive the Whittle indices for a nonlinear objective function, specially, convex or concave function.

# Publications

- [1] KRIOUILE Saad, LARRANAGA Maialen, and ASSAAD Mohamad. Whittle index policy for multichannel scheduling in queueing systems. In : 2019 IEEE International Symposium on Information Theory (ISIT). IEEE, 2019. p. 2524-2528.
- [2] MAATOUK Ali, ASSAAD Mohamad, et al. Asymptotically Optimal Scheduling Policy For Minimizing The Age of Information. In : 2020 IEEE International Symposium on Information Theory (ISIT). IEEE, 2020. p. 1747-1752: The first and the second authors contributed equally.
- [3] MAATOUK Ali, KRIOUILE Saad, ASSAAD Mohamad, et al. On The Optimality of The Whittle's Index Policy For Minimizing The Age of Information, accepted in IEEE Transactions on Wireless Communications: The first and the second authors contributed equally.
- [4] MAATOUK Ali, KRIOUILE Saad, ASSAAD Mohamad, et al. The age of incorrect information: A new performance metric for status updates, accepted in IEEE/ACM Transactions on Networking.
- [5] KRIOUILE Saad, ASSAAD Mohamad, MAATOUK Ali. On The Global Optimality of The Whittle's Index Policy For Minimizing The Age of Information submitted in IEEE Transactions on Information Theory.
- [6] KRIOUILE Saad, ASSAAD Mohamad, and LARRANAGA Maialen. Asymptotically Optimal Delay-aware Scheduling in Queueing Systems submitted in IEEE Journal of Communications and Networks.



---

# Bibliography

- [1] “Number of internet of things (iot) connected devices worldwide in 2018, 2025 and 2030,” <https://www.statista.com/statistics/802690/worldwide-connected-devices-by-access-technology/>, accessed:Feb 19 2020.
- [2] “Little law,” <https://en.wikipedia.org/wiki/Little-law/>.
- [3] S. Kaul, R. Yates, and M. Gruteser, “On piggybacking in vehicular networks,” in *2011 IEEE Global Telecommunications Conference-GLOBECOM 2011*. IEEE, 2011, pp. 1–5.
- [4] —, “Real-time status: How often should one update?” in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 2731–2735.
- [5] S. K. Kaul, R. D. Yates, and M. Gruteser, “Status updates through queues,” in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2012, pp. 1–6.
- [6] R. D. Yates and S. Kaul, “Real-time status updating: Multiple sources,” in *2012 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2012, pp. 2666–2670.
- [7] C. Q. Wu, X. Lin, D. Yu, W. Xu, and L. Li, “End-to-end delay minimization for scientific workflows in clouds under budget constraint,” *IEEE Transactions on Cloud Computing*, vol. 3, no. 2, pp. 169–181, 2014.
- [8] C.-y. Liu, B. Fu, and H.-J. Huang, “Delay minimization and priority scheduling in wireless mesh networks,” *Wireless networks*, vol. 20, no. 7, pp. 1955–1965, 2014.
- [9] J. Liu, Y. Mao, J. Zhang, and K. B. Letaief, “Delay-optimal computation task scheduling for mobile-edge computing systems,” in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 1451–1455.
- [10] J. He, L. Cai, P. Cheng, and J. Pan, “Delay minimization for data dissemination in large-scale vanets with buses and taxis,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 1939–1950, 2015.
- [11] B. Nazir and H. Hasbullah, “Dynamic sleep scheduling for minimizing delay in wireless sensor network,” in *2011 Saudi International Electronics, Communications and Photonics Conference (SIECPC)*. IEEE, 2011, pp. 1–5.

- 
- [12] Y. Zhang, P. Du, J. Wang, T. Ba, R. Ding, and N. Xin, “Resource scheduling for delay minimization in multi-server cellular edge computing systems,” *IEEE Access*, vol. 7, pp. 86 265–86 273, 2019.
- [13] D. Rajan, A. Sabharwal, and B. Aazhang, “Delay-bounded packet scheduling of bursty traffic over wireless channels,” *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 125–144, 2004.
- [14] M. Andrews, A. Fernandez Anta, L. Zhang, and W. Zhao, “Routing and scheduling for energy and delay minimization in the powerdown model,” *Networks*, vol. 61, no. 3, pp. 226–237, 2013.
- [15] Y. Cui and V. K. Lau, “Distributive stochastic learning for delay-optimal ofdma power and subband allocation,” *IEEE transactions on signal processing*, vol. 58, no. 9, pp. 4848–4858, 2010.
- [16] Y. Cui, V. K. Lau, R. Wang, H. Huang, and S. Zhang, “A survey on delay-aware resource control for wireless systems—large deviation theory, stochastic lyapunov drift, and distributed stochastic learning,” *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1677–1701, 2012.
- [17] Q. Zheng, K. Zheng, H. Zhang, and V. C. Leung, “Delay-optimal virtualized radio resource scheduling in software-defined vehicular networks via stochastic learning,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 7857–7867, 2016.
- [18] R. Wang and V. K. Lau, “Delay-aware two-hop cooperative relay communications via approximate mdp and stochastic learning,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7645–7670, 2013.
- [19] Y. Ruan, W. Wang, Z. Zhang, and V. K. Lau, “Delay-aware massive random access for machine-type communications via hierarchical stochastic learning,” in *Communications (ICC), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1–6.
- [20] J. Li, M. Peng, A. Cheng, Y. Yu, and C. Wang, “Resource allocation optimization for delay-sensitive traffic in fronthaul constrained cloud radio access networks,” *IEEE Systems Journal*, vol. 11, no. 4, pp. 2267–2278, 2014.
- [21] L. Lei, Y. Kuang, N. Cheng, X. Shen, Z. Zhong, and C. Lin, “Delay-optimal dynamic mode selection and resource allocation in device-to-device communications—part ii: Practical algorithm,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3491–3505, 2015.
- [22] I. Bettesh and S. Shamai, “Optimal power and rate control for minimal average delay: The single-user case,” *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4115–4141, 2006.
- [23] K. Liu and Q. Zhao, “Indexability of restless bandit problems and optimality of whittle index for dynamic multichannel access,” *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5547–5567, 2010.

- 
- [24] P. Ansell, K. D. Glazebrook, J. Niño-Mora, and M. O’Keeffe, “Whittle index policy for a multi-class queueing system with convex holding costs,” *Mathematical Methods of Operations Research*, vol. 57, no. 1, pp. 21–39, 2003.
- [25] S. Aalto, P. Lassila, and P. Osti, “Whittle index approach to size-aware scheduling with time-varying channels,” in *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 2015, pp. 57–69.
- [26] N. Akbarzadeh and A. Mahajan, “Restless bandits with controlled restarts: Indexability and computation of whittle index,” in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 7294–7300.
- [27] K. E. Avrachenkov and V. S. Borkar, “Whittle index policy for crawling ephemeral content,” *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 446–455, 2016.
- [28] V. S. Borkar, “Whittle index for partially observed binary markov decision processes,” *IEEE Transactions on Automatic Control*, vol. 62, no. 12, pp. 6614–6618, 2017.
- [29] F. Iannello, O. Simeone, and U. Spagnolini, “Optimality of myopic scheduling and whittle indexability for energy harvesting sensors,” in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2012, pp. 1–6.
- [30] M. Larrañaga, M. Assaad, A. Destounis, and G. S. Paschos, “Asymptotically optimal pilot allocation over markovian fading channels,” *IEEE Transactions on Information Theory*, 2017.
- [31] R. Meshram, D. Manjunath, and A. Gopalan, “On the whittle index for restless multiarmed hidden markov bandits,” *IEEE Transactions on Automatic Control*, vol. 63, no. 9, pp. 3046–3053, 2018.
- [32] W. Ouyang, A. Eryilmaz, and N. B. Shroff, “Downlink scheduling over markovian fading channels,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1801–1812, 2015.
- [33] B. Sombabu, A. Mate, D. Manjunath, and S. Moharir, “Whittle index for aoi-aware scheduling,” in *2020 International Conference on COMMunication Systems & NETWORKS (COMSNETS)*. IEEE, 2020, pp. 630–633.
- [34] V. Tripathi and E. Modiano, “A whittle index approach to minimizing functions of age of information,” in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 1160–1167.
- [35] J. Wang, X. Ren, Y. Mo, and L. Shi, “Whittle index policy for dynamic multichannel allocation in remote state estimation,” *IEEE Transactions on Automatic Control*, vol. 65, no. 2, pp. 591–603, 2019.
- [36] P. Whittle, “Restless bandits: Activity allocation in a changing world,” *Journal of applied probability*, vol. 25, no. A, pp. 287–298, 1988.

- 
- [37] R. R. Weber and G. Weiss, “On an index policy for restless bandits,” *Journal of Applied Probability*, vol. 27, no. 3, pp. 637–648, 1990.
- [38] M. Larrañaga, U. Ayesta, and I. M. Verloop, “Dynamic control of birth-and-death restless bandits: Application to resource-allocation problems.” *IEEE/ACM Trans. Netw.*, vol. 24, no. 6, pp. 3812–3825, 2016.
- [39] J. A. Van Mieghem, “Dynamic scheduling with convex delay costs: The generalized  $c|\mu$  rule,” *The Annals of Applied Probability*, pp. 809–833, 1995.
- [40] A. Mandelbaum and A. L. Stolyar, “Scheduling flexible servers with convex delay costs: Heavy-traffic optimality of the generalized  $c\mu$ -rule,” *Operations Research*, vol. 52, no. 6, pp. 836–855, 2004.
- [41] S. K. Kaul and R. D. Yates, “Age of information: Updates with priority,” in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2644–2648.
- [42] A. M. Bedewy, Y. Sun, and N. B. Shroff, “The age of information in multihop networks,” *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1248–1257, 2019.
- [43] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, “Scheduling policies for minimizing age of information in broadcast wireless networks,” *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2637–2650, 2018.
- [44] Y.-P. Hsu, E. Modiano, and L. Duan, “Age of information: Design and analysis of optimal scheduling algorithms,” in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 561–565.
- [45] —, “Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals,” *IEEE Transactions on Mobile Computing*, 2019.
- [46] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, “Age-optimal updates of multiple information flows,” in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2018, pp. 136–141.
- [47] B. T. Bacinoglu and E. Uysal-Biyikoglu, “Scheduling status updates to minimize age of information with an energy harvesting sensor,” in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 1122–1126.
- [48] E. T. Ceran, D. Gündüz, and A. György, “Average age of information with hybrid arq under a resource constraint,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 3, pp. 1900–1913, 2019.
- [49] —, “Reinforcement learning to minimize age of information with an energy harvesting sensor with harq and sensing cost,” in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2019, pp. 656–661.

- 
- [50] H. B. Beytur and E. Uysal, “Age minimization of multiple flows using reinforcement learning,” in *2019 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2019, pp. 339–343.
- [51] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, “Minimizing age of information in vehicular networks,” in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. IEEE, 2011, pp. 350–358.
- [52] J. Gittins, K. Glazebrook, and R. Weber, *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.
- [53] C. H. Papadimitriou and J. N. Tsitsiklis, “The complexity of optimal queuing network control,” *Mathematics of Operations Research*, vol. 24, no. 2, pp. 293–305, 1999.
- [54] S. M. Ross, *Introduction to stochastic dynamic programming*. Academic press, 2014.
- [55] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [56] P. K. Dutta, “What do discounted optima converge to?: A theory of discount rate asymptotics in economic models,” *Journal of Economic Theory*, vol. 55, no. 1, pp. 64–94, 1991.
- [57] M. Larrañaga, “Dynamic control of stochastic and fluid resource-sharing systems,” Ph.D. dissertation, 2015.
- [58] K. Liu, R. Weber, and Q. Zhao, “Indexability and whittle index for restless bandit problems involving reset processes,” in *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE, 2011, pp. 7690–7696.
- [59] T. G. Kurtz, “Strong approximation theorems for density dependent markov chains,” *Stochastic Processes and their Applications*, vol. 6, no. 3, pp. 223–240, 1978.



# Appendix A

## Appendix: Methodology

### A.1 Proof of Proposition 3.1

We consider the Bellman Equation (3.8). By summing the RHS and the LHS of Equation (3.8), for all  $i$  we obtain:

$$\sum_{k=1}^N [V_k(s_k) + \theta_k] = \sum_{k=1}^N \min_{a_k} \{C_k(s_k, a_k) + \sum_{s'_k} Pr(s'_k | s_k, a_k) V_k(s'_k)\} \quad (\text{A.1})$$

$$= \min_{\mathbf{a}} \left\{ \sum_{k=1}^N [C_k(s_k, a_k) + \sum_{s'_k} Pr(s'_k | s_k, a_k) V_k(s'_k)] \right\}, \quad (\text{A.2})$$

where  $\mathbf{a} = (a_1, \dots, a_N)$ . We also have that:

$$Pr(\mathbf{s}' | \mathbf{s}, \mathbf{a}) = \sum_{s'_k} Pr(\mathbf{s}' | \mathbf{s}, \mathbf{a}, a'_k) Pr(s'_k | \mathbf{s}, \mathbf{a}) = \sum_{s'_k} Pr(\mathbf{s}' | \mathbf{s}, \mathbf{a}, s'_k) Pr(s'_k | s_k, a_k), \quad (\text{A.3})$$

for all  $\mathbf{s} = (s_1, \dots, s_N)$  and  $\mathbf{s}' = (s'_1, \dots, s'_N)$ . Since  $Pr(s_k | \mathbf{s}, \mathbf{a})$  only depends on the decision taken with respect to user  $i$ , we obtain:

$$\sum_{k=1}^N \sum_{s'_k} Pr(s'_k | s_k, a_k) V_k(s'_k) = \sum_{k=1}^N \sum_{\mathbf{s}'} \sum_{s'_k} Pr(\mathbf{s}' | \mathbf{s}, \mathbf{a}, s'_k) Pr(s'_k | s_k, a_k) V_k(s'_k) \quad (\text{A.4})$$

$$= \sum_{\mathbf{s}'} Pr(\mathbf{s}' | \mathbf{s}, \mathbf{a}) \sum_{k=1}^N V_k(s'_k) \quad (\text{A.5})$$

From the previous equations we obtain:

$$\sum_{k=1}^N V_k(s_k) + \sum_{k=1}^N \theta_k = \min_{\mathbf{a}} \left[ \sum_{k=1}^N C_k(s_k, a_k) + \sum_{k=1}^N \sum_{s'_k} Pr(s'_k | s_k, a_k) V_k(s'_k) \right] \quad (\text{A.6})$$

$$= \min_{\mathbf{a}} \left[ \sum_{k=1}^N C_k(s_k, a_k) + \sum_{\mathbf{s}'} Pr(\mathbf{s}' | \mathbf{s}, \mathbf{a}) \sum_{k=1}^N V_k(s'_k) \right] \quad (\text{A.7})$$

According to Theorem 2.1 Chapter 2, [54], it exists a unique function  $\bar{V}$  and a constant  $\theta$  that resolves Equation (3.7). Subsequently, since we have found a bounded function  $\sum_{k=1}^N V_k(s_k)$ , and a constant  $\sum_{k=1}^N \theta_k$  that satisfy also Equation (3.7), then  $\bar{V}(s) = \sum_{k=1}^N V_k(s_k)$  and  $\theta = \sum_{k=1}^N \theta_k$ . This is equivalent to find for each user the decision that minimizes the right hand side of each individual Bellman equation. This concludes the proof.

## A.2 Proof of Proposition 3.2

Before proving the proposition, we give two useful lemmas.

**Lemma A.1.** *Considering  $a_{j-1}, a_j, a_{j+1}$  and  $b_{j-1}, b_j, b_{j+1}$ , such that  $b_{j-1} < b_j < b_{j+1}$ .*

1. If  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$   
Then:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.8})$$

2. If  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$  Then:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.9})$$

3. If  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}}$   
Then:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.10})$$

4. If  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}}$  Then:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.11})$$

5. If  $\frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$  Then:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.12})$$

6. If  $\frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$  Then:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.13})$$

*Proof.* We will just prove the first case. For the other cases, the proof is similar.

First case:  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \implies \frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$ .

For the LHS inequality:

$$\frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} = \frac{a_{j+1} - a_j}{b_{j+1} - b_{j-1}} + \frac{a_j - a_{j-1}}{b_{j+1} - b_{j-1}} \quad (\text{A.14})$$

$$\geq \frac{(a_j - a_{j-1})(b_{j+1} - b_j)}{(b_j - b_{j-1})(b_{j+1} - b_{j-1})} + \frac{a_j - a_{j-1}}{b_{j+1} - b_{j-1}} \quad (\text{A.15})$$

The inequality above comes from the fact that  $b_{j-1} < b_j < b_{j+1}$  and  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$

Then

$$\frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_j - a_{j-1}}{b_j - b_{j-1}} \left[ \frac{b_{j+1} - b_j + b_j - b_{j-1}}{b_{j+1} - b_{j-1}} \right] \quad (\text{A.16})$$

$$= \frac{a_j - a_{j-1}}{b_j - b_{j-1}} \quad (\text{A.17})$$

For the RHS inequality:

$$\frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} = \frac{a_{j+1} - a_j}{b_{j+1} - b_{j-1}} + \frac{a_j - a_{j-1}}{b_{j+1} - b_{j-1}} \quad (\text{A.18})$$

$$\leq \frac{a_{j+1} - a_j}{b_{j+1} - b_{j-1}} + \frac{(a_{j+1} - a_j)(b_j - b_{j-1})}{(b_{j+1} - b_j)(b_{j+1} - b_{j-1})} \quad (\text{A.19})$$

where the above inequality comes from the fact that  $b_{j-1} < b_j < b_{j+1}$  and  $\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j}$

Then

$$\frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \left[ \frac{b_{j+1} - b_j + b_j - b_{j-1}}{b_{j+1} - b_{j-1}} \right] \quad (\text{A.20})$$

$$= \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \quad (\text{A.21})$$

■

**Lemma A.2.** *The largest minimizer at step  $j$  in algorithm 1 satisfies  $n_j = \min\{k : b_k = b_{n_j}\}$*

*Proof.* We consider  $i$  such that  $b_i = b_{n_j}$  and we prove that  $n_j \leq i$ :

By construction of  $n_j$ ,  $b_{n_{j-1}} \neq b_{n_j}$  and  $n_{j-1} < n_j$ . Hence, by increase of  $b_k$ ,  $b_{n_j} \geq b_{n_{j-1}}$ .

Therefore  $b_i = b_{n_j} > b_{n_{j-1}}$ , and  $i > n_{j-1}$ . Consequently, according to definition of  $n_j$ :

$$\frac{a_{n_j} - a_{n_{j-1}}}{b_{n_j} - b_{n_{j-1}}} \leq \frac{a_i - a_{n_{j-1}}}{b_i - b_{n_{j-1}}} \quad (\text{A.22})$$

$$\frac{a_{n_j} - a_{n_{j-1}}}{b_{n_j} - b_{n_{j-1}}} \leq \frac{a_i - a_{n_{j-1}}}{b_{n_j} - b_{n_{j-1}}} \quad (\text{A.23})$$

This implies that  $a_{n_j} \leq a_i$ .

If  $i < n_j$ , as  $b_i = b_{n_j}$ , then  $a_i < a_{n_j}$  which contradicts with  $a_{n_j} \leq a_i$ .

Therefore  $n_j \leq i$ . This concludes the proof. ■

We start by the indexability then we show that the expressions given by algorithm 1 are effectively the Whittle's indices.

- Indexability:

We consider  $n_1$  and  $n_2$  the optimal thresholds of the problem (3.17) when the lagrangian parameter  $W$  is equal to  $W_1$  and  $W_2$  respectively, such that  $W_1 < W_2$ . To that extent, we show that  $n_1$  is less than  $n_2$ . In fact, establishing the aforementioned result is sufficient to show the indexability since by proving it, we can say that the set of states for which the optimal decision is passive action when  $W = W_1$ , is included in the set of states for which the optimal decision is passive action when  $W = W_2$ , specifically  $[0, n_1] \subseteq [0, n_2]$ . Subsequently:  $D(W_1) \subseteq D(W_2)$ .

In order to prove that, we just need to demonstrate that  $b_{n_1} \leq b_{n_2}$  since  $n_1 \leq n_2$  is equivalent to  $b_{n_1} \leq b_{n_2}$ , due to the increase of  $b_n$  with  $n$ .

As  $n_1$  and  $n_2$  are the minimizers of Equation (4.6) when  $W = W_1$  and  $W = W_2$  respectively, then:

$$a_{n_1} - W_1 b_{n_1} \leq a_{n_2} - W_1 b_{n_2} \quad (\text{A.24})$$

$$a_{n_1} - W_2 b_{n_1} \geq a_{n_2} - W_2 b_{n_2} \quad (\text{A.25})$$

This implies:

$$W_2(b_{n_1} - b_{n_2}) \leq a_{n_1} - a_{n_2} \leq W_1(b_{n_1} - b_{n_2}) \quad (\text{A.26})$$

Therefore:  $(W_2 - W_1)(b_{n_1} - b_{n_2}) \leq 0$ . Since  $W_2 - W_1 > 0$ , thus  $b_{n_1} \leq b_{n_2}$ . Consequently,  $n_1 \leq n_2$ .

Thereby, we conclude the indexability.

- Whittle's index expressions:

For the Whittle's index expressions, we should demonstrate that, for  $k \in ]n_{j-1}, n_j]$ ,  $W_j = \min\{W, k \in D(W)\}$ .

For that, we prove first that for  $W < W_j$ ,  $k \notin D(W)$ .

If  $k > n_{j-1}$ ,  $W < W_j$ , and  $b_k \neq b_{n_{j-1}}$ , then  $W < W_j \leq \frac{a_k - a_{n_{j-1}}}{b_k - b_{n_{j-1}}}$ . Therefore,  $a_k - b_k W > a_{n_{j-1}} - b_{n_{j-1}} W$ .

If  $k > n_{j-1}$ ,  $W < W_j$  and  $b_k = b_{n_{j-1}}$ , given that  $a_k > a_{n_{j-1}}$ , then  $a_k - b_k W > a_{n_{j-1}} - b_{n_{j-1}} W$ .

Hence we have proved that, for  $W < W_j$  and  $k > n_{j-1}$ ,  $a_k - b_k W > a_{n_{j-1}} - b_{n_{j-1}} W$ . That means for  $W < W_j$ , the optimal threshold is  $n_{j-1}$  or even less. Therefore, for  $k \in ]n_{j-1}, n_j]$ , the optimal action is the active one, i.e.  $k \notin D(W)$ .

We still have to prove that  $k \in D(W_j)$ .

For that, we prove that the optimal threshold is at least  $n_j$  when  $W = W_j$ . In other words, for all  $k < n_j$ ,  $a_k - b_k W_j \geq a_{n_j} - b_{n_j} W_j$ . We demonstrate this result by induction in  $j$ :

–  $j = 0$

By definition,  $W_0 \leq \frac{a_k - a_{-1}}{b_k} \forall k \geq 0$ . Furthermore, as  $b_n$  is increasing with  $n$ , then  $b_k \leq b_{n_0}$  for  $0 \leq k < n_0$ . However, according to Lemma A.2,  $b_k$  is necessarily strictly less than  $b_{n_0}$ . Thus, by using Lemma A.1 (fourth case), we can deduce that  $\frac{a_{n_0} - a_k}{b_{n_0} - b_k} \leq W_0$ . That means, as  $b_{-1} = 0$ , we have for  $k \in [-1, n_0[$ ,

$\frac{a_{n_0}-a_k}{b_{n_0}-b_k} \leq W_0$ , which implies that  $a_k - b_k W_0 \geq a_{n_0} - b_{n_0} W_0$ .

- We suppose at step  $j$ ,  $a_k - b_k W_j \geq a_{n_j} - b_{n_j} W_j$  i.e.  $\frac{a_{n_j}-a_k}{b_{n_j}-b_k} \leq W_j$  for  $k < n_j$  (this remains true since  $b_k < b_{n_j}$  according to Lemma A.2). We show that  $a_k - b_k W_{j+1} \geq a_{n_{j+1}} - b_{n_{j+1}} W_{j+1}$  for  $k < n_{j+1}$ , i.e.:  
When  $n_j \leq k < n_{j+1}$ , then if  $b_k \neq b_{n_j}$ ,  $\frac{a_k-a_{n_j}}{b_k-b_{n_j}} \geq W_{j+1}$ . Thus, by using Lemma A.1 (fourth case), we get  $\frac{a_{n_{j+1}}-a_k}{b_{n_{j+1}}-b_k} \leq W_{j+1}$  ( $b_{n_j} < b_k < b_{n_{j+1}}$ ). If  $b_k = b_{n_j}$ ,  $\frac{a_{n_{j+1}}-a_k}{b_{n_{j+1}}-b_k} = \frac{a_{n_{j+1}}-a_k}{b_{n_{j+1}}-b_{n_j}} \leq \frac{a_{n_{j+1}}-a_{n_j}}{b_{n_{j+1}}-b_{n_j}} = W_{j+1}$  since  $a_k \geq a_{n_j}$ .  
When  $k < n_j$ , we have that  $\frac{a_{n_j}-a_k}{b_{n_j}-b_k} \leq W_j$  (induction assumption). By definition of  $n_j$  in Algorithm 1, we have  $W_j < \frac{a_{n_{j+1}}-a_{n_{j-1}}}{b_{n_{j+1}}-b_{n_{j-1}}}$ . Then according to Lemma A.1 (third case),  $W_j \leq W_{j+1}$  ( $b_{n_{j-1}} < b_{n_j} < b_{n_{j+1}}$ ). Therefore  $\frac{a_{n_j}-a_k}{b_{n_j}-b_k} \leq W_{j+1}$  and by using again Lemma A.1 (first case),  $\frac{a_{n_{j+1}}-a_k}{b_{n_{j+1}}-b_k} \leq W_{j+1}$ . Therefore, for all  $k \leq n_{j+1}$ ,  $a_k - b_k W_{j+1} \geq a_{n_{j+1}} - b_{n_{j+1}} W_{j+1}$ .

As consequence, we have proved by induction that at any step  $j$ , for  $k < n_j$ ,  $a_k - b_k W_j \geq a_{n_j} - b_{n_j} W_j$ .

Then when  $W = W_j$ , the optimal threshold is at least  $n_j$ . This means that if  $k \in ]n_{j-1}, n_j]$ , then  $k$  is surely less or equal than the optimal threshold when  $W = W_j$ , which implies that the optimal decision at state  $k$  is passive action, i.e.  $k \in D(W_j)$ . Combining the two results for  $k \in ]n_{j-1}, n_j]$ :

- For  $W < W_j$ ,  $k \notin D(W)$ .
- $k \in D(W_j)$ .

Then  $W_j = \min\{W, k \in D(W)\}$ . This concludes the proof.



## Appendix B

# Appendix: Whittle's index policy for minimizing the Delay in Queuing system

### B.1 Proof of Lemma 4.2

We first prove that  $C(\cdot, \cdot)$  is submodular. That is,  $(C(q+1, 1) - C(q+1, 0)) - (C(q, 1) - C(q, 0)) = a(q+1) + W - a(q+1) - (aq + W - aq) = 0 \leq 0$ . The latter is obtained by substituting the values of  $C(q', s)$  for  $s \in \{0, 1\}$  and  $q' \in \{q, q+1\}$ . In order to prove that  $\sum_{q'} Pr(q'|q, s)V(q')$  is submodular, we distinguish between two cases:

Case 1)  $q < R$ , then:

$$\begin{aligned}
 & \sum_{q'} Pr(q'|q+1, 1)V(q') - \sum_{q'} Pr(q'|q+1, 0)V(q') \\
 &= \sum_{q'=0} Pr(A=q')V(q') - \sum_{q'=q+1} Pr(A=q'-q-1)V(q') \\
 &= \sum_{q'=0} Pr(A=q')V(q') - \sum_{q'=q} Pr(A=q'-q)V(q'+1) \\
 &\leq \sum_{q'=0} Pr(A=q')V(q') - \sum_{q'=q} Pr(A=q'-q)V(q') \\
 &= \sum_{q'} Pr(q'|q, 1)V(q') - \sum_{q'} Pr(q'|q, 0)V(q') \tag{B.1}
 \end{aligned}$$

The inequality follows from the fact that  $V(\cdot)$  is increasing. This concludes the proof for  $q < R$ .

Case 2)  $q \geq R$ , then:

$$\begin{aligned}
 \sum_{q'} Pr(q'|q+1, 1)V(q') - \sum_{q'} Pr(q'|q+1, 0)V(q') \\
 &= \sum_{q'} Pr(A = q' - q - 1 + R)V(q') - \sum_{q'} Pr(A = q' - q - 1)V(q') \\
 &= \sum_{q'=q+1-R} Pr(A = q' - q - 1 + R)V(q') - \sum_{q'=q+1} Pr(A = q' - q - 1)V(q') \\
 &= \sum_{q'=q} Pr(A = q' - q)V(q' - R + 1) - \sum_{q'=q} Pr(A = q' - q)V(q' + 1)
 \end{aligned} \tag{B.2}$$

Moreover, we have:

$$\sum_{q'} Pr(q'|q, 1)V(q') - \sum_{q'} Pr(q'|q, 0)V(q') = \sum_{q'=q} Pr(A = q' - q)V(q' - R) - \sum_{q'=q} Pr(A = q' - q)V(q'). \tag{B.3}$$

Subtracting Equation (B.2) and (B.3) (i.e., (B.2)-(B.3)), we obtain:

$$\sum_{q'=q} Pr(A = q' - q)[(V(q' - R + 1) - V(q' - R)) - (V(q' + 1) - V(q'))] \leq 0, \tag{B.4}$$

which follows from the R-convexity of  $V(\cdot)$ . Therefore,  $\sum_{q'} Pr(q'|q, s)V(q')$  is submodular.

## B.2 Proof of Proposition 4.1

When  $i < L$ :

1)  $j \leq n$ :

Since  $j \leq n$ , the optimal decision is to stay idle, that means if  $A$  denotes the number of arrival packets, in the next time slot the number of packets will be  $i = j + A$  with  $A \leq R - 1$ , then  $A = i - j$ . Therefore, the probability to transit from state  $j$  to  $i$  is the probability that  $A = i - j$ , which is exactly  $\pi_{i-j}$ .

2)  $j > n$ :

The optimal decision in this case is to transmit. As  $j \leq L < R$ , then all  $j$  packets will be transmitted. Taking into account the  $A$  arrival packets, then the new state for the next time slot will be  $i = A$ . This explains that the probability to transit from state  $j$  to  $i$  is the probability that  $A$  is equal to  $i$  which is equal to  $\pi_i$ .

When  $i = L$ :

1)  $j \leq n$ :

The optimal decision is the passive action. Then,  $A$  arriving packets are added to the  $j$  packets present in the queue. At the next time slot, the number of packets is  $j + A$ . According to Equation (4.1), since we cannot exceed the buffer length  $L$ , we reach the state  $L$  if  $j + A \geq L$ . Since  $A \leq R - 1$ , then the probability of this event or equivalently the probability to transit from state  $j$  to state  $L$  is  $Pr(L - j \leq A \leq R - 1) = \sum_{k=L-j}^{R-1} Pr(A = k) = (R - L + j)\pi_{L-j} = (R - L + j)\rho$ .

2)  $j > n$ :

The optimal decision is the active action. Subsequently, the next state is 0 as  $j < R$ . Thus to reach the state  $L$ , the arrival packet number  $A$  must be in the set  $[L, R - 1]$ . Therefore, the probability to transit from 0 to  $L$  is  $Pr(L \leq A \leq R - 1) = \sum_{k=L}^{R-1} Pr(A = k) = (R - L)\pi_L = (R - L)\rho$ .

### B.3 Proof of Proposition 4.2

We have that:

$$u(i) = \sum_{j=0}^n p^n(j, i)u(j) + \sum_{j=n+1}^L p^n(j, i)u(j) \quad (\text{B.5})$$

We distinguish between two cases:  $n < L$  and  $n = L$ . We analyze each case separately.

1.  $n < L$ :

We first give the expression of  $u(i)$  when  $i < L$  based on Proposition 4.1:

$$u(i) = \sum_{j=0}^n \pi_{i-j}u(j) + \sum_{j=n+1}^L \pi_i u(j) \quad (\text{B.6})$$

By definition of  $\pi$  given in Definition 4.3, we have that:

$$u(i) = \sum_{j=0}^{\min(i, n)} \rho u(j) + \sum_{j=n+1}^L \rho u(j) \quad (\text{B.7})$$

Now, in order to prove Proposition 4.2 for this case, we will distinguish between tree sub-cases:

- a)  $n + 1 \leq i \leq L - 1$
- b)  $0 \leq i \leq n$
- c)  $i = L$

a) Proof of  $u(i) = \rho$  for  $n + 1 \leq i \leq L - 1$ :

We have  $\min(i, n) = n$ , then:

$$u(i) = \sum_{j=0}^n \rho u(j) + \sum_{j=n+1}^L \rho u(j) \quad (\text{B.8})$$

Knowing that  $\sum_0^L u(j) = 1$ , thus  $\sum_0^n \rho u(j) + \sum_{n+1}^L \rho u(j) = \rho$ . Hence,  $u(i) = \rho$ .

b) Proof of  $u(i) = \rho(1 - \rho)^{n-i}$  for  $0 \leq i \leq n$ :

We prove this result by induction, i.e., we start by proving that the statement  $P(i) = \{u(i) = \rho(1 - \rho)^{n-i}\}$  holds for  $i = n$ , then we show that it holds for  $i - 1$ , if  $P(i - 1)$  is true.

- $i = n$ :

We have that:  $u(n) = \sum_0^n \rho u(j) + \sum_{n+1}^L \rho u(j) = \rho = \rho(1 - \rho)^{n-n}$ . Thereby,  $P(n)$  is true.

- $P(i) \Rightarrow P(i-1)$ :

We have that  $\min(i-1, n) = i-1$ , then:

$$\begin{aligned} u(i-1) &= \sum_0^{i-1} \rho u(j) + \sum_{n+1}^L \rho u(j) \\ u(i-1) &= \sum_0^i \rho u(j) + \sum_{n+1}^L \rho u(j) - \rho u(i) \\ u(i-1) &= u(i) - \rho u(i) \end{aligned} \tag{B.9}$$

By induction assumption, we have that  $u(i) = \rho(1-\rho)^{n-i}$ . To that extent, replacing the expression of  $u(i)$  in (B.9), we obtain:

$$u(i-1) = (1-\rho)u(i) = \rho(1-\rho)^{n-(i-1)}$$

That concludes the proof.

As for  $i = L$ ,  $u(L)$  is nothing but the subtraction of the  $\sum_{j=0}^{L-1} u(j)$  from 1. By doing so, we get:

$$u(L) = (1-\rho)^{n+1} - (L-n-1)\rho$$

2.  $n = L$ :

$$u(i) = \sum_{j=0}^L p^L(j, i) u(j) \tag{B.10}$$

For  $i \leq L-1$ :

According to Proposition 4.1, we have:

$$u(i) = \sum_{j=0}^L \pi_{i-j} u(j) \tag{B.11}$$

By definition of  $\pi$ , we get:

$$u(i) = \sum_0^i \rho u(j) \tag{B.12}$$

We prove by induction that for  $0 \leq i < L$ ,  $u(i) = 0$

We have  $u(0) = \rho u(0) = 0$ .

We suppose that  $u(j) = 0$  for all  $0 \leq j \leq i$ , then:

$$u(i+1) = \sum_0^{i+1} \rho u(j) \tag{B.13}$$

$$= \sum_0^i \rho u(j) + \rho u(i+1) \tag{B.14}$$

$$= 0 + \rho u(i+1) \tag{B.15}$$

$$u(i+1) = 0 \tag{B.16}$$

Then, for all  $i \in [0, L-1]$ ,  $u(i) = 0$ .

Since  $\sum_{j=0}^L u(j) = 1$ , we have  $u(L) = 1 - \sum_{j=0}^{L-1} u(j) = 1 - 0 = 1$ .

This ends the proof.

## B.4 Proof of Proposition 4.3

When  $n \in [0, L - 1]$ , we have  $a_n - a_{n-1} = a[\rho[(L - n) - (L + R)(1 - \rho)^n] + 1]$ . To that extent, we denote by  $f(n)$  the function:  $a[\rho[(L - n) - (L + R)(1 - \rho)^n] + 1]$  and we show that  $f(\cdot)$  is positive for  $n \in [0, L]$ . To that end, we give the second derivative of  $f(\cdot)$ :

$$f''(n) = a[-\rho(\ln(1 - \rho))^2(L + R)(1 - \rho)^n] \quad (\text{B.17})$$

It is clear from the above equation that  $f''(\cdot)$  is non positive. Hence,  $f(\cdot)$  is concave function with  $n$ . That is, for all  $n \in [0, L]$ ,  $f(n) \geq \min\{f(0), f(L)\}$ . Thereby, our task will be to demonstrate that  $f(0)$  and  $f(L)$  are both positive. In fact,  $f(0) = 0 \geq 0$ . While for  $n = L$ , it requires more technical analysis to establish the desired result. Computing  $f(L)$ , we get:

$$f(L) = a[1 - \rho(L + R)(1 - \rho)^L]$$

We have  $f(L) - f(0) = f(L) = \sum_{n=0}^{L-1} f(n+1) - f(n)$ . To that extent, we give the expression of  $v(n) = f(n+1) - f(n)$ , i.e.:

$$v(n) = a[\rho[-1 + \rho(L + R)(1 - \rho)^n]] \quad (\text{B.18})$$

knowing that  $v(n)$  is lower bounded by  $a[\rho[-1 + \rho(L + R)(1 - \rho)^L]]$  for  $n \in [0, L - 1]$ , then:

$$\begin{aligned} f(L) &\geq \sum_{n=0}^{L-1} a[\rho[-1 + \rho(L + R)(1 - \rho)^L]] \\ &= a[L\rho[-1 + \rho(L + R)(1 - \rho)^L]] \end{aligned} \quad (\text{B.19})$$

Therefore:

$$f(L) = a[1 - \rho(L + R)(1 - \rho)^L] \geq a[-L\rho[1 - \rho(L + R)(1 - \rho)^L]] \quad (\text{B.20})$$

From the above inequality,  $1 - \rho(L + R)(1 - \rho)^L$  should be positive otherwise, we will have a non positive term higher than a positive term. Consequently,  $f(L)$  is positive. Providing that  $f(n) \geq \min\{f(0), f(L)\}$  for  $n \in [0, L]$ , then  $f(n) \geq 0$ . Hence for  $n \in [0, L - 1]$ ,

$$a_n \geq a_{n-1} \quad (\text{B.21})$$

We still have to show that  $a_L - a_{L-1} \geq 0$ . In fact:

$$a_L - a_{L-1} = a[R - (L + R)(1 - \rho)^L] = a[R(1 - \rho(L + R)(1 - \rho)^L)] \geq 0 \quad (\text{B.22})$$

Thus, combining the two results (B.21) and (B.22), we end up with the desired result.

## B.5 Proof of Theorem 4.3

In order to prove this theorem, we introduce the following useful lemmas.

**Lemma B.1.**  $x_{n,n-1}$  is strictly increasing with  $n$

*Proof.* We have for all  $n \in [0, L - 1]$ :

$$x_{n+1,n} - x_{n,n-1} = W(n+1) - W(n) = \frac{a\rho^2(L-n)}{\rho(1-\rho)^{n+1}} > 0$$

That concludes the proof. ■

**Lemma B.2.** *If for any  $k \in [0, L - 1]$ , we have that:  $b_{k-1} < b_k < b_{k+1}$  and  $\frac{a_k - a_{k-1}}{b_k - b_{k-1}} < \frac{a_{k+1} - a_k}{b_{k+1} - b_k}$ .*

*Then for any  $k \in [0, L - 1]$ , we have for each  $k < s \leq L$ :*

$$\frac{a_s - a_{k-1}}{b_s - b_{k-1}} > \frac{a_k - a_{k-1}}{b_k - b_{k-1}} \quad (\text{B.23})$$

*Proof.* We fix certain  $k \in [0, L - 1]$ , we prove the result by induction:  
for  $s = k + 1$

$$\frac{a_{k+1} - a_{k-1}}{b_{k+1} - b_{k-1}} = \frac{a_{k+1} - a_{k-1} - a_k + a_k}{b_{k+1} - b_{k-1}} \quad (\text{B.24})$$

$$= \frac{a_{k+1} - a_k}{b_{k+1} - b_{k-1}} + \frac{a_k - a_{k-1}}{b_{k+1} - b_{k-1}} \quad (\text{B.25})$$

$$> \frac{(a_k - a_{k-1})(b_{k+1} - b_k)}{(b_k - b_{k-1})(b_{k+1} - b_{k-1})} + \frac{(a_k - a_{k-1})(b_k - b_{k-1})}{(b_k - b_{k-1})(b_{k+1} - b_{k-1})} \quad (\text{B.26})$$

where the strict inequality comes from the lemma's assumptions. Therefore, we have that:

$$\frac{a_{k+1} - a_{k-1}}{b_{k+1} - b_{k-1}} > \frac{a_k - a_{k-1}}{b_k - b_{k-1}} \left[ \frac{b_{k+1} - b_k}{b_{k+1} - b_{k-1}} + \frac{b_k - b_{k-1}}{b_{k+1} - b_{k-1}} \right] \quad (\text{B.27})$$

$$= \frac{a_k - a_{k-1}}{b_k - b_{k-1}} \quad (\text{B.28})$$

By induction, we consider that the inequality (B.23) is true for certain  $s$  strictly higher than  $k$ . The inequality below is then verified for  $s + 1$ :

$$\frac{a_{s+1} - a_{k-1}}{b_{s+1} - b_{k-1}} = \frac{a_{s+1} - a_{k-1} - a_s + a_s}{b_{s+1} - b_{k-1}} \quad (\text{B.29})$$

$$= \frac{a_{s+1} - a_s}{b_{s+1} - b_{k-1}} + \frac{a_s - a_{k-1}}{b_{s+1} - b_{k-1}} \quad (\text{B.30})$$

$$> \frac{(a_k - a_{k-1})(b_{s+1} - b_s)}{(b_k - b_{k-1})(b_{s+1} - b_{k-1})} + \frac{(a_k - a_{k-1})(b_s - b_{k-1})}{(b_k - b_{k-1})(b_{s+1} - b_{k-1})} \quad (\text{B.31})$$

$$= \frac{a_k - a_{k-1}}{b_k - b_{k-1}} \left[ \frac{b_{s+1} - b_s}{b_{s+1} - b_{k-1}} + \frac{b_s - b_{k-1}}{b_{s+1} - b_{k-1}} \right] \quad (\text{B.32})$$

$$= \frac{a_k - a_{k-1}}{b_k - b_{k-1}}. \quad (\text{B.33})$$

So the inequality is also true for  $s + 1$ . This concludes the proof of the lemma. ■

Referring to Algorithm 1 that allows us to obtain the Whittle indices, we denote by  $j$  the step  $j$  described in the algorithm.

According to the same algorithm, to establish that  $x_{j,j-1}$  is the Whittle's index at the

state  $j$ , we need to prove that for all  $n \in [j+1, L]$ ,  $\frac{a_n - a_{j-1}}{b_n - b_{j-1}} > \frac{a_j - a_{j-1}}{b_j - b_{j-1}}$ . Indeed, using Lemma B.1,  $W(j) < W(j+1) < \dots < W(L)$ . Therefore, for all  $k \in [j, L-1]$ ,  $\frac{a_k - a_{k-1}}{b_k - b_{k-1}} < \frac{a_{k+1} - a_k}{b_{k+1} - b_k}$ . Hence, according to Lemma B.2, for all  $n \in [j+1, L]$ ,  $\frac{a_n - a_{j-1}}{b_n - b_{j-1}} > \frac{a_j - a_{j-1}}{b_j - b_{j-1}}$ . Thus, the minimizer of  $\frac{a_n - a_{j-1}}{b_n - b_{j-1}}$  at step  $j$  is  $j$ . As consequence, the Whittle index of state  $j$  according to Algorithm 2 is effectively  $W(j) = \frac{a_j - a_{j-1}}{b_j - b_{j-1}} = x_{j,j-1}$ .

## B.6 Proof of Proposition 4.4

1)  $n \leq R_k - 1$ :

We start first by giving an useful lemma

**Lemma B.3.** *For all  $0 \leq i \leq p \leq R_k - 1$ ,*

$$V^p(R_k + i, W) - V^p(i, W) = a_k R_k + W$$

*Proof.* We decompose the discounted cost  $V^p(i + R_k, W)$  in the cost incurred at first time slot plus the discounted cost starting at the next time slot. At state  $i + R_k$ , the decision taken is to transmit since  $i + R_k \geq R_k > p$ , and at state  $i$ , the decision taken is passive action since  $i \leq p$ , hence,

$$V^p(i + R_k, W) = a_k(i + R_k) + W + \rho_k \beta \sum_{j=0}^{R_k-1} V^p(i + j, W) \quad (\text{B.34})$$

$$V^p(i, W) = a_k i + \rho_k \beta \sum_{j=0}^{R_k-1} V^p(i + j, W) \quad (\text{B.35})$$

Subtracting the second term from the first term,

$$V^p(i + R_k, W) - V^p(i, W) = a_k R_k + W \quad (\text{B.36})$$

□

We have:

$$C_1^m(n, W) = a_k n + W + \beta \rho_k \sum_{i=0}^{R_k-1} V^n(i, W) \quad (\text{B.37})$$

$$C_0^m(n, W) = a_k n + \beta \rho_k \sum_{i=0}^{R_k-1} V^n(i + n, W) \quad (\text{B.38})$$

That means,

$$C_1^n(n, W) - C_0^n(n, W) = W + \beta\rho_k \sum_{i=0}^{R_k-1} V^n(i, W) - \beta\rho_k \sum_{i=0}^{R_k-1} V^n(i+n, W) \quad (\text{B.39})$$

$$= W + \beta\rho_k \sum_{i=0}^{R_k-1} V^n(i, W) - \beta\rho_k \sum_{i=n}^{R_k-1+n} V^n(i, W) \quad (\text{B.40})$$

$$= W + \beta\rho_k \sum_{i=0}^{n-1} V^n(i, W) - \beta\rho_k \sum_{i=R_k}^{R_k-1+n} V^n(i, W) \quad (\text{B.41})$$

$$= W + \beta\rho_k \sum_{i=0}^{n-1} V^n(i, W) - \beta\rho_k \sum_{i=0}^{n-1} V^n(i+R_k, W) \quad (\text{B.42})$$

Applying the Lemma B.3,

$$C_1^n(n, W) - C_0^n(n, W) = W - \beta\rho_k n(a_k R_k + W) \quad (\text{B.43})$$

$$C_1^n(n, W) - C_0^n(n, W) = W(1 - \beta\rho_k n) - \beta n a_k \quad (\text{B.44})$$

2)  $n \geq R_k$ :

We consider a given threshold  $n \geq R_k$ , i.e., for states less than  $n$ , we don't transmit otherwise we transmit. At state  $n$  if we decide to transmit then the next possible states are  $n - R_k + i$  ( $i$  varies from 0 to  $R_k - 1$ ) with the probability to reach each state is  $\rho_k$ , hence, we have,

$$C_1^n(n, W) = a_k n + W + \beta\rho_k \sum_{i=0}^{R_k-1} V^n(n - R_k + i, W)$$

At state  $n + i - R_k$ , since  $n - R_k + i < n$  then, the decision taken is passive action ( $n$  is threshold), thus if we decompose again  $V^n(n - R_k + i, W)$ ,  $V^n(n + i - R_k, W) = a_k(n + i - R_k) + \beta\rho_k \sum_{j=0}^{R_k-1} V^n(n - R_k + i + j, W)$ , Replacing  $V^n(n + i - R_k, W)$  by its value,

$$C_1^n(n, W) = a_k n + W + \beta\rho_k \sum_{i=0}^{R_k-1} [a_k(n + i - R_k) + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n - R_k + i + j, W)] \quad (\text{B.45})$$

$$\begin{aligned} C_1^n(n, W) = & a_k n + W + \beta\rho_k \sum_{i=1}^{R_k-1} [a_k(n + i - R_k) + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n - R_k + i + j, W)] \\ & + \beta\rho_k [a_k(n - R_k) + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n - R_k + j, W)] \end{aligned} \quad (\text{B.46})$$

We know that,

$$C_1^n(n, W) = a_k n + W + \beta\rho_k \sum_{j=0}^{R_k-1} V^n(n - R_k + j, W) \quad (\text{B.47})$$

Hence,

$$C_1^n(n, W) = a_k n + W + \beta \rho_k \sum_{i=1}^{R_k-1} [a_k(n+i-R_k) + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)] + \beta \rho_k C_1^n(n, W) - \rho_k \beta W - a_k \beta \quad (B.48)$$

That means,

$$C_1^n(n, W) = \frac{1}{1 - \rho_k \beta} [a_k n + W + \beta \rho_k \sum_{i=1}^{R_k-1} [a_k(n+i-R_k) + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)] - \rho_k \beta W - a_k \beta] \quad (B.49)$$

$$C_1^n(n, W) = \frac{1}{1 - \rho_k \beta} [a_k n + W + \beta \rho_k \sum_{i=1}^{R_k-1} a_k(n+i-R_k) + \rho_k^2 \beta^2 \sum_{i=1}^{R_k-1} \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)] - \rho_k \beta W - a_k \beta \quad (B.50)$$

At state  $n$  if we decide to not transmit then the next possible states are  $n+i$  with the probability to reach each state is  $\rho_k$ . Thus, we have,  $C_0^n(n, W) = a_k n + \beta \rho_k \sum_{i=0}^{R_k-1} V^n(n+i, W)$  At state  $n+i$  for  $i > 0$ , since  $n+i > n$  then, the decision taken is active action ( $n$  is threshold), thus if we decompose again  $V^n(n+i, W)$ ,  $V^n(n+i, W) = a_k(n+i) + W + \beta \rho_k \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)$ , Replacing  $V^n(n+i, W)$  when  $i > 0$  by its value,

$$C_0^n(n, W) = a_k n + \beta \rho_k \sum_{i=1}^{R_k-1} [a_k(n+i) + W + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)] + \beta \rho_k V^n(n, W) \quad (B.51)$$

However  $V^n(n, W)$  is no more than  $C_0^n(n, W)$ , hence,

$$C_0^n(n, W) = a_k n + \beta \rho_k \sum_{i=1}^{R_k-1} [a_k(n+i) + W + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)] + \beta \rho_k C_0^n(n, W) \quad (B.52)$$

That means,

$$C_0^n(n, W) = \frac{1}{1 - \rho_k \beta} [a_k n + \beta \rho_k \sum_{i=1}^{R_k-1} [a_k(n+i) + W + \rho_k \beta \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)]] \quad (B.53)$$

$$C_0^n(n, W) = \frac{1}{1 - \rho_k \beta} [a_k n + \beta \rho_k \sum_{i=1}^{R_k-1} [a_k(n+i) + W] + \rho_k^2 \beta^2 \sum_{i=1}^{R_k-1} \sum_{j=0}^{R_k-1} V^n(n-R_k+i+j, W)] \quad (B.54)$$

Then,

$$\begin{aligned} (C_1^n(n, W) - C_0^n(n, W))(1 - \rho_k \beta) &= W - \beta \rho_k \sum_{i=1}^{R_k-1} [a_k R_k + W] - \rho_k \beta W - a_k \beta \\ &= W - \beta(R_k - 1)a_k - \beta \rho_k (R_k - 1)W - \rho_k \beta W - a_k \beta \\ &= W - \beta a_k R_k - \beta W \\ &= W(1 - \beta) - \beta a_k R_k \end{aligned} \quad (B.55)$$

Hence,  $C_1^n(n, W) - C_0^n(n, W) = \frac{W(1-\beta) - \beta a_k R_k}{1 - \rho_k \beta}$

## B.7 Proof of Proposition 4.5

As the function  $TO(V^n)$  is submodular, then for all  $q_i^k \leq n$ , we have  $C_0^n(q_i^k, W) - C_1^n(q_i^k, W) \leq C_0^n(n, W) - C_1^n(n, W) = 0$ , and for all  $q_i^k > n$ , we have  $C_0^n(q_i^k, W) - C_1^n(q_i^k, W) \geq C_0^n(n, W) - C_1^n(n, W) = 0$ . That means  $n$  is indeed an optimal threshold.

## B.8 Proof of Proposition 4.6

1)  $n \leq R_k - 1$ :

According to Proposition 4.4 for  $n \leq R_k - 1$ ,  $C_1^n(n, W) - C_0^n(n, W) = g_k(n, W) = W(1 - n\beta\rho_k) - a_k\beta n$ . Knowing that  $g_k(n, W) = 0 \Leftrightarrow W = \frac{\beta a_k R_k n}{R_k - \beta n}$ . That means, for  $W = \frac{\beta a_k R_k n}{R_k - \beta n}$ ,  $C_1^n(n, W) - C_0^n(n, W) = 0$ . Hence, using Proposition 4.5,  $n$  is indeed an optimal threshold.

2)  $n \geq R_k$ :

According to Proposition 4.4 for  $n \geq R_k$ ,  $C_1^n(n, W) - C_0^n(n, W) = g_k(n, W) = \frac{W(1-\beta) - a_k\beta R_k}{1 - \rho_k \beta}$ . Knowing that  $g_k(n, W) = 0 \Leftrightarrow W = \frac{\beta a_k R_k}{1 - \beta}$ . Hence,  $C_1^n(n, W) = C_0^n(n, W) \Leftrightarrow W = \frac{\beta a_k R_k}{1 - \beta}$ . Applying Proposition 4.5, the threshold  $n$  is indeed an optimal solution when  $W = \frac{\beta a_k R_k}{1 - \beta}$ . That is true for all  $n \geq R_k$ , which concludes the proof.

## B.9 Proof of Theorem 4.4

Before proving the Theorem, we give an useful proposition.

**Proposition B.1.** *If  $W < \frac{a_k R_k \beta}{1 - \beta}$ , then the optimal threshold is surely finite.*

*Proof.* We consider that the infinite threshold policy denoted by  $\infty$  is the optimal solution, and its respective Value function is  $V^\infty$ .

**Lemma B.4.** *For all  $q_i^k$ ,  $V^\infty(q_i^k + R_k, W) - V^\infty(q_i^k, W) \geq \frac{a_k R_k}{1 - \beta}$ .*

*Proof.* Under the infinite threshold, since the decision taken for all states is passive action, then we have for all  $q_i^k \geq 0$ ,

$$V^\infty(q_i^k + R_k, W) = a_k(q_i^k + R_k) + \beta\rho_k \sum_{i=1}^{R_k-1} V^\infty(q_i^k + R_k + i, W)$$

and

$$V^\infty(q_i^k, W) = a_k q_i^k + \beta\rho_k \sum_{i=1}^{R_k-1} V^\infty(q_i^k + i, W)$$

Hence,

$$V^\infty(q_i^k + R_k, W) - V^\infty(q_i^k, W) = a_k R_k + \beta\rho_k \sum_{i=1}^{R_k-1} [V^\infty(q_i^k + R_k + i, W) - V^\infty(q_i^k + i, W)] \tag{B.56}$$

Due to the  $R_k$ -convexity of  $V^\infty$ , then  $V^\infty(q_i^k + R_k + i, W) - V^\infty(q_i^k + i, W) > V^\infty(q_i^k + R_k, W) - V^\infty(q_i^k, W)$ . Therefore,  $V^\infty(q_i^k + R_k, W) - V^\infty(q_i^k, W) \geq a_k R_k + \beta[V^\infty(q_i^k + R_k, W) - V^\infty(q_i^k, W)]$ . That implies  $V^\infty(q_i^k + R_k, W) - V^\infty(q_i^k, W) \geq \frac{a_k R_k}{1-\beta}$ .  $\square$

We have the difference between  $C_1^\infty(q_i^k, W)$  and  $C_0^\infty(q_i^k, W)$  for  $q_i^k \geq R_k$ ,

$$C_1^\infty(q_i^k, W) - C_0^\infty(q_i^k, W) = W + \beta \rho_k \sum_{i=0}^{R_k-1} [V^\infty(q_i^k + i - R_k, W) - V^\infty(q_i^k + i, W)] \quad (\text{B.57})$$

Applying Lemma B.4

$$C_1^\infty(q_i^k, W) - C_0^\infty(q_i^k, W) \leq W + \beta \rho_k \sum_{i=0}^{R_k-1} \left[-\frac{a_k R_k}{1-\beta}\right] \quad (\text{B.58})$$

According to Proposition's assumption, we have that  $W < \frac{a_k R_k \beta}{1-\beta}$ , therefore,

$$C_1^\infty(q_i^k, W) - C_0^\infty(q_i^k, W) < \frac{a_k R_k \beta}{1-\beta} - \frac{a_k R_k \beta}{1-\beta} = 0 \quad (\text{B.59})$$

Thus, at state  $q_i^k$ , the optimal decision is to transmit which contradict with the fact that the threshold is infinite. That concludes the proof of the present proposition.  $\square$

1)  $0 \leq n \leq R_k - 1$ :

We fix a state  $n$  less than  $R_k - 1$ . For  $W < \frac{a_k R_k n \beta}{R_k - \beta n}$ , as  $g_k(\cdot)$  is strictly increasing with  $W$ , then  $g_k(n, W)$  is strictly less than  $g_k(n, \frac{a_k R_k n \beta}{R_k - \beta n}) = 0$ . As  $g_k$  is decreasing with  $n$ , then for all  $q_i^k \geq n$ ,  $g_k(q_i^k, W) \leq g_k(n, W) < 0$ . That implies for all  $q_i^k \geq n$ ,  $q_i^k$  can not be threshold (otherwise there exists  $q_i^k \geq n$  such that  $g_k(q_i^k, W) \geq 0$ ). Moreover, since  $W < \frac{a_k R_k n \beta}{R_k - \beta n} \leq \frac{a_k R_k \beta}{1-\beta}$ , then according to Proposition B.1, the optimal threshold must be finite. Thereby, surely the optimal threshold is strictly less than  $n$ . That means at state  $n$  the optimal decision is the active action. Hence, when  $W < \frac{a_k R_k n \beta}{R_k - \beta n}$ ,  $n \notin D(W)$ . Applying Proposition 4.6, for  $W = \frac{a_k R_k n \beta}{R_k - \beta n}$ , the threshold  $n$  is an optimal solution. Then  $n \in D(W)$ . Consequently, we conclude the result for the first case.

2)  $n \geq R_k$ : For this case, we need to prove that for all  $n \geq R_k$ ,  $\frac{a_k R_k \beta}{1-\beta} = \min\{W, n \in D(W)\}$ . In other words, for all  $W < \frac{a_k R_k \beta}{1-\beta}$ ,  $n \notin D(W)$ , and  $n \in D(\frac{a_k R_k \beta}{1-\beta})$ . When  $W < \frac{a_k R_k \beta}{1-\beta}$ , as  $g_k(\cdot)$  is strictly increasing with  $W$ , then  $g_k(n, W)$  is strictly less than  $g_k(n, \frac{a_k R_k \beta}{1-\beta}) = 0$  for all  $n \geq R_k$ . That implies for all  $n \geq R_k$ ,  $n$  can not be the optimal threshold. Moreover according to Proposition B.1, the optimal threshold must be finite. Hence, the optimal threshold is surely strictly less than  $R_k$ . That means at state  $n \geq R_k$ , the optimal decision is the active action, i.e., when  $W < \frac{a_k R_k \beta}{1-\beta}$ ,  $n \notin D(W)$ . Applying Proposition 4.6, for  $W = \frac{a_k R_k \beta}{1-\beta}$ , the threshold  $n$  is an optimal solution, then  $n \in D(W)$ . Hence, we conclude the result.

## B.10 Proof of Theorem 4.5

In order to prove this result, we need to show that the order from the biggest Whittle index to the smallest one is the same.

First of all, we give the order of the Whittle indices given by Theorem 4.4, when  $\beta > 1 - \frac{\min\{a_j R_j\}}{\max\{a_j R_j^2\}}$ .

Due to the indexability of all classes, it's obvious that the Whittle index is increasing with  $n$  for a given class  $k$ . Moreover, considering any two classes  $k$  and  $m$ , then, the Whittle index of any state  $n_k$  in class  $k$  and greater than  $R_k$  is larger than the Whittle index of any state  $n_m$  in class  $m$  less than  $R_m$ . In fact we have  $\beta > 1 - \frac{\min\{a_j R_j\}}{\max\{a_j R_j^2\}}$ , that means  $1 - \beta < \frac{\min\{a_j R_j\}}{\max\{a_j R_j^2\}}$ . Hence,  $\frac{1}{1-\beta} > \frac{\max\{a_j R_j^2\}}{\min\{a_j R_j\}}$ . Then,  $W_k(n_k) = \frac{a_k R_k}{1-\beta} > \frac{a_k R_k \max\{a_j R_j^2\}}{\min\{a_j R_j\}} \geq \max\{a_j R_j^2\} \geq a_m R_m^2 \geq a_m R_m (R_m - 1) \geq \beta a_m R_m (R_m - 1) \geq \frac{\beta a_m R_m (R_m - 1)}{R_m - \beta(R_m - 1)}$  (because  $R_m - \beta(R_m - 1) \geq 1$ )  $= W_m(R_m - 1) \geq W_m(n_m)$ .

For the new form of Whittle index where we get rid of  $\beta$  for states greater than maximum transmission rate, the order is not affected for the states less than  $R_k - 1$ , since the Whittle indices are the same. For the states greater than  $R_k$ , the Whittle index  $a_k R_k \max\{a_j R_j^2\}$  is higher than  $a_k R_k a_m R_m (R_m - 1) \geq a_m R_m (R_m - 1) \geq \frac{\beta a_m R_m (R_m - 1)}{R_m - \beta(R_m - 1)} = W_m(R_m - 1)$ . Furthermore, the order between the Whittle indices greater than the transmission rate doesn't change since we just multiply by a constant which is  $\frac{1-\beta}{\max\{a_j R_j^2\}}$  to go from  $\frac{a_k R_k}{1-\beta}$  to  $a_k R_k \max\{a_j R_j^2\}$ . Hence, the order of these new expressions of Whittle indices is not affected. That means, by adopting this policy, we end up with the same Whittle index policy in Theorem 4.4.

## B.11 Proof of Proposition 4.7

In order to prove this proposition, we distinguish between two types of classes:

- 1) Class  $k$  in which  $W$  is different from all  $W_i^k$ .
- 2) Class  $k$  such that there exists a given state  $j$  that satisfies  $W_j^k = W$ .

First type of classes: For the class  $k$  in which  $W$  is different from all  $W_i^k$ , we prove that the optimal threshold verifies  $l_k(W) = l_k = \arg \max_i \{W_i^k | W_i^k \leq W\} = \arg \max_i \{W_i^k | W_i^k < W\}$ . First we have  $\arg \max_i \{W_i^k | W_i^k \leq W\} = \arg \max_i \{W_i^k | W_i^k < W\}$  since  $W_i^k$  is different from  $W$  for all state  $i$ . For state  $i$  less than  $l_k$ , given that  $W_i^k$  is increasing with  $i$ , then  $W_i^k \leq W_{l_k}^k < W$ . Hence, due to the indexability of the class,  $D(W_i^k) \subseteq D(W)$ , which implies that the optimal decision at state  $i$  is the passive action. For the state  $i$  strictly greater than  $l_k$ , by definition of  $l_k$ ,  $W_i^k$  must be strictly greater than  $W$  since  $l_k$  is the integer that gives the highest Whittle index less than  $W$ . Then, according to the definition of Whittle index,  $W < \min\{W, i \in D(W)\}$ , that means  $W \notin \{W, i \in D(W)\}$ . Therefore  $i \notin D(W)$ . Thus, the optimal decision at state  $i > l_k$  is the active decision. Hence  $l_k = \arg \max_i \{W_i^k | W_i^k \leq W\} = \arg \max_i \{W_i^k | W_i^k < W\}$  is effectively the optimal threshold  $l_k(W)$ .

Now, we tackle the case when there exists  $j$ ,  $W_j^k = W$ :

We know that according to Theorem 4.3,  $W_j^k = x_{j,j-1}^k$  which is the point for which if  $W = x_{j,j-1}^k$ , we have  $\sum_{q=0}^L a_k u_k^j(q)q - W \sum_{q=0}^j u_k^j(q) = \sum_{q=0}^L a_k u_k^{j-1}(q)q - W \sum_{q=0}^{j-1} u_k^{j-1}(q)$ . That means, according to Equation (4.47), for  $W = x_{j,j-1}^k$ , if  $j$  is a minimizer of this equation ( $j$  is the optimal threshold), then  $j-1$  is also a minimizer of this equation. Due to the indexability of the class  $k$ , for all states less or equal than  $j$ , the optimal decision

is to stay passive. Besides, according to the definition of Whittle index, for all states strictly higher than  $j$ , the optimal decision is to be active. Then,  $j$  is indeed an optimal threshold, so as for  $j - 1$ .

Hence, the optimal threshold can be either  $j$  or  $j - 1$ .

In fact, since  $W_0^k < \dots < W_{j-1}^k < W_j^k = W$ , then  $j = \arg \max_i \{W_i^k | W_i^k \leq W\}$ , and  $j - 1 = \arg \max_i \{W_i^k | W_i^k < W\}$ .

This proves the proposition.

## B.12 Proof of Proposition 4.8

From optimization theory, it is known that the optimal solution of the dual problem is less or equal than the primal problem's solution when the constraint is satisfied, i.e:

$$\max_W \min_{\phi \in \Phi} f(W, \phi) \leq \min_{\phi \in \Phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} a_k q_i^k(t) \mid \mathbf{q}(0), \phi \right] \quad (\text{B.60})$$

As the optimal solution for a fixed  $W$  is a threshold-based policy, we use the steady state form and the expression of the LHS of the above inequality becomes:

$$\max_W \min_{\phi} f(W, \phi) = \max_W \left\{ \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} \left[ \min_{l_k \in [-1, L]} \left\{ \sum_{q=0}^L a_k u_k^{l_k}(q) q - W \sum_{q=0}^{l_k} u_k^{l_k}(q) \right\} \right] + W(1 - \alpha)N \right\} \quad (\text{B.61})$$

with  $\phi$  being the threshold policy that corresponds to  $l(W)$  computed using Proposition 4.7 for a fixed  $W$ . For  $W^*$  that satisfies the constraint with equality (i.e.  $\alpha N = \sum_{k=1}^K \gamma_k N \sum_{i=l_{k+1}(W^*)}^L u_k^{l_k(W^*)}(i)$ , which is in fact true for all  $N$ , and then we can get rid of  $N$ ), we have:  $\sum_{k=1}^K \sum_{i=1}^{\gamma_k N} [-W \sum_{q=0}^{l_k(W^*)} u_k^{l_k(W^*)}(q)] + W(1 - \alpha)N = \sum_{k=1}^K \gamma_k N [-W(1 - \sum_{i=l_{k+1}(W^*)}^L u_k^{l_k(W^*)}(i))] + W(1 - \alpha)N = -NW + \sum_{k=1}^K \gamma_k N W \sum_{i=l_{k+1}(W^*)}^L u_k^{l_k(W^*)}(i) + W(1 - \alpha)N = -NW + \alpha N + WN - \alpha N = 0$ . Hence, we get:

$$\begin{aligned} \min_{\phi} f(W^*, \phi) &= f(W^*, l(W^*)) = \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} \left[ \sum_{q=0}^L a_k u_k^{l_k(W^*)}(q) q \right] \\ &= \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} a_k q_i^k(t) \mid \mathbf{q}(0), l(W^*) \right] \quad (\text{B.62}) \end{aligned}$$

Therefore, we obtain a threshold vector  $l(W^*)$  that gives us a solution for the constrained relaxed problem (primal problem) that satisfies the constraint (4.3). Moreover, according to the inequality (B.60), we have that for all policy  $\phi$  that satisfies the constraint and

belong to  $\Phi$ :

$$\begin{aligned}
 f(W^*, l(W^*)) &= \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} \left[ \sum_{q=0}^L a_k u_k^{l_k(W^*)}(q) q \right] = \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} a_k q_i^k(t) \mid \mathbf{q}(0), l(W^*) \right] \\
 &= \min_{\phi} f(W^*, \phi) \\
 &\leq \max_W \min_{\phi} f(W, \phi) \\
 &\leq \min_{\phi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{\gamma_k N} a_k q_i^k(t) \mid \mathbf{q}(0), \phi \right].
 \end{aligned} \tag{B.63}$$

We deduce that the solution of the relaxed problem is of type threshold-based policy  $l(W^*)$  with  $W^*$  satisfies  $\alpha = \sum_{k=1}^K \gamma_k \sum_{i=l_{k+1}(W^*)}^L u_k^{l_k(W^*)}(i)$ .

## B.13 Proof of Proposition 4.9

We define the following order relation in  $\mathbb{R}^K$  such that for any two vectors  $l^1$  and  $l^2$ ,  $l^1 \leq l^2 \iff$  for each element of vector of index  $k$ , we have  $l_k^1 \leq l_k^2$ . Recall that according to Proposition 4.7, we can directly deduce that for  $W_1 \leq W_2$   $l(W_1) \leq l(W_2)$ .

Without loss of generality, when  $W \in \mathbb{R}^+$ , the corresponding set of threshold vectors  $l(W)$  is perfectly ordered. Then,  $\sum_{k=1}^K \gamma_k \sum_{i=l_k(W)+1}^L u_k^{l_k(W)}(i)$  is strictly decreasing in  $l(W)$ , and take discrete values from 1 to 0. According to Proposition 4.7, we have for each class  $k$  and state  $i$ , if  $W = W_i^k$  then there is two possible optimal thresholds vectors  $l^1(W)$  and  $l^2(W)$  with  $l^1(W) < l^2(W)$ . Hence we can deduce that there exists a class  $m$  and state  $p$  such that  $\sum_{k=1}^K \gamma_k \sum_{i=l_k^1(W_p^m)+1}^L u_k^{l_k^1(W_p^m)}(i) \geq \alpha$  and  $\sum_{k=1}^K \gamma_k \sum_{i=l_k^2(W_p^m)+1}^L u_k^{l_k^2(W_p^m)}(i) \leq \alpha$ .

According to Proposition 4.7, when  $W = W_p^m$ ,  $l_m(W_p^m) = l_m^2(W_p^m)$  and  $l_m^1(W_p^m) = l_m^2(W_p^m) - 1 = l_m(W_p^m) - 1$  can be both the optimal thresholds for class  $m$ . As for the other classes,  $l_k^1(W_p^m) = l_k^2(W_p^m) = l_k(W_p^m)$ .

If we force  $W^*$  to be equal to  $W_p^m$ , the optimal threshold vector can be either  $l^1(W_p^m)$  or  $l^2(W_p^m)$ , then we can introduce some randomization between the two policies. In other words, we use the threshold policy  $l^1(W_p^m)$  with probability  $\theta$  and  $l^2(W_p^m)$  with probability  $1 - \theta$ . The new stationary distribution for the class  $m$  is then a linear combination of these two threshold policies  $l_m(W_p^m)$  and  $l_m(W_p^m) - 1$ :  $u_m^* = \theta u_m^{l_m(W_p^m)} + (1 - \theta) u_m^{l_m(W_p^m) - 1}$ . Hence, in the class  $m$ , at state strictly less than  $l_m(W_p^m)$ , the queues will not transmit, whereas in a state strictly greater than  $l_m(W_p^m)$ , they will transmit with probability one. If the queues are in state  $l_m(W_p^m)$ , they will transmit with probability  $\frac{(1 - \theta) u_m^{l_m(W_p^m) - 1}(l_m(W_p^m))}{\theta u_m^{l_m(W_p^m)}(l_m(W_p^m)) + (1 - \theta) u_m^{l_m(W_p^m) - 1}(l_m(W_p^m))} = \frac{(1 - \theta) u_m^{l_m(W_p^m) - 1}(l_m(W_p^m))}{u_m^*(l_m(W_p^m))}$ . Since the probability to be in this state  $l_m(W_p^m)$  is  $u_m^*(l_m(W_p^m))$ , the proportion of time that the queues will be in active mode is:

$$\sum_{k \neq m} \sum_{i=l_k(W_p^m)+1}^L \gamma_k u_k^{l_k(W_p^m)}(i) + \sum_{i=l_m(W_p^m)+1}^L \gamma_m u_m^*(i) + (1 - \theta) \gamma_m u_m^{l_m(W_p^m) - 1}(l_m(W_p^m))$$

When  $\theta = 0$ , the threshold policy is  $l_m(W_p^m) - 1$  and the total average time in active mode is higher than  $\alpha$ . When  $\theta = 1$ , the threshold policy is  $l_m(W_p^m)$  and the total average time in active mode is less than  $\alpha$ .

Given that  $\sum_{k \neq m} \sum_{i=l_k(W_p^m)+1}^L \gamma_k u_k^{l_k(W_p^m)}(i) + \sum_{i=l_m(W_p^m)+1}^L \gamma_m u_m^*(i) + (1-\theta) \gamma_m u_m^{l_m(W_p^m)-1}(l_m(W_p^m))$  is continuous with  $\theta$ , then there exists  $\theta^*$  which verifies the equality. Hence, for  $W^* = W_p^m$ , we get a threshold policy for all classes except for class  $m$  where the optimal solution is a linear combination of two threshold policies. Moreover for a given randomized parameter  $\theta^*$ , the constraint (4.3) is satisfied with equality:

$$\alpha = \sum_{k \neq m} \sum_{i=l_k(W_p^m)+1}^L \gamma_k u_k^{l_k(W_p^m)}(i) + \sum_{i=l_m(W_p^m)+1}^L \gamma_m u_m^*(i) + (1 - \theta^*) \gamma_m u_m^{l_m(W_p^m)-1}(l_m(W_p^m))$$

## B.14 Proof of Proposition 4.10

We derive the eigenvalues of  $\mathbf{Q}$ .

The matrix  $\mathbf{Q}$  is of the form:

$$\begin{bmatrix} Q_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & Q_2 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & & \\ A_1 & A_2 & \cdots & Q_m & \cdots & A_{K-1} & A_K \\ \vdots & & & \ddots & & \vdots & \\ 0 & 0 & \cdots & \cdots & \cdots & Q_{K-1} & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & Q_K \end{bmatrix} \quad (\text{B.64})$$

The characteristic polynomial of  $\mathbf{Q}$  is the product of the characteristic polynomial of each matrix  $Q_k$ :

$$\chi_{\mathbf{Q}}(\lambda) = \prod_{k=1}^K \chi_{Q_k}(\lambda) \quad (\text{B.65})$$

Therefore, the set of  $\mathbf{Q}$ 's eigenvalues denoted by  $\text{Sp}(\mathbf{Q})$  is composed by the eigenvalues of the matrices  $Q_k$ . Specifically:  $\text{Sp}(\mathbf{Q}) = \bigcup_k \text{Sp}(Q_k)$ .

1) The case  $k \neq m$ :

$$Q_k =$$

$$\begin{array}{c}
 0 \\
 1 \\
 \vdots \\
 l_k - 2 \\
 l_k - 1 \\
 l_k + 1 \\
 \vdots \\
 L - 1 \\
 L
 \end{array}
 \begin{pmatrix}
 0 & 1 & \cdots & l_k - 2 & l_k - 1 & l_k + 1 & l_k + 2 & \cdots & L - 1 & L \\
 \rho_k & 0 & \cdots & 0 & 0 & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
 \rho_k & \rho_k & \ddots & 0 & 0 & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
 \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \rho_k & \rho_k & \vdots & \vdots \\
 \vdots & & \ddots & \rho_k & 0 & \vdots & & & \vdots & \vdots \\
 \rho_k & \cdots & \cdots & \rho_k & \rho_k & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
 \vdots & 0 & 0 & \vdots & \vdots & \vdots & 0 & 0 & \vdots & \vdots \\
 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
 -l_k \rho_k & (1 - l_k) \rho_k & \cdots & -2 \rho_k & -\rho_k & -l_k \rho_k & \cdots & \cdots & -l_k \rho_k & -l_k \rho_k
 \end{pmatrix}$$

After computations and some algebraic manipulations, we get,  $\chi_{Q_k}(\lambda) = (-\lambda)^L$

2)The case  $k = m$ :

$$Q_m =$$

$$\begin{array}{c}
 0 \\
 1 \\
 \vdots \\
 l_m - 2 \\
 l_m - 1 \\
 l_m + 1 \\
 \vdots \\
 L - 1 \\
 L
 \end{array}
 \begin{pmatrix}
 0 & 1 & \cdots & l_m - 2 & l_m - 1 & l_m + 1 & l_m + 2 & \cdots & L - 1 & L \\
 \rho_m & 0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
 \rho_m & \rho_m & \ddots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
 \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & 0 & 0 & \vdots & \vdots \\
 \vdots & & \ddots & \rho_m & 0 & \vdots & & & \vdots & \vdots \\
 \rho_m & \cdots & \cdots & \rho_m & \rho_m & 0 & \cdots & \cdots & 0 & 0 \\
 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
 \vdots & 0 & 0 & \vdots & \vdots & \vdots & 0 & 0 & \vdots & \vdots \\
 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
 -l_m \rho_m & (1 - l_m) \rho_m & \cdots & -2 \rho_m & -\rho_m & 0 & \cdots & \cdots & 0 & 0
 \end{pmatrix}$$

After computations and some algebraic manipulations, we get:

$$\chi_{Q_m}(\lambda) = (-\lambda)^{L-l_m} (\rho_m - \lambda)^{l_m}$$

For  $k \neq m$ ,  $Q_k$  has only 0 as eigenvalue.

For  $k = m$ ,  $\chi_{Q_m}(\lambda) = 0 \Leftrightarrow \lambda = 0$  or  $\lambda = \rho_m$ . Hence,  $Q_m$  has two eigenvalues: 0 and  $\rho_m$  which are strictly less than 1.

Consequently, in both cases, the norms of all eigenvalues of  $Q_k$  are strictly less than 1. Hence, for  $\lambda \in \text{Sp}(\mathbf{Q}) \Rightarrow |\lambda| < 1$ .

## B.15 Proof of Lemma 4.6

We take  $0 < \epsilon < \mu$ ,  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$ , i.e. there exists  $T_0$  such that for all  $t \geq T_0$ ,  $\|\mathbf{z}(t) - \mathbf{z}^*\| \leq \epsilon$ . Hence:

$$P_x\left(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu\right) \leq P_x\left(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| + \|\mathbf{z}(t) - \mathbf{z}^*\| \geq \mu\right) \quad (\text{B.66})$$

$$\leq P_x\left(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu - \epsilon\right) \quad (\text{B.67})$$

$$\leq P_x\left(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu - \epsilon\right) \quad (\text{B.68})$$

Using Proposition 4.11, there exists  $s_1$  and  $s_2$  such that:

$$P_x\left(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu - \epsilon\right) \leq s_1 \exp(-Ns_2). \quad (\text{B.69})$$

Therefore:

$$P_x\left(\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu\right) \leq s_1 \exp(-Ns_2). \quad (\text{B.70})$$

## B.16 Proof of Proposition 4.12

We recall that  $\mathbf{Z}^N(t)$  represents the proportion vector at time  $t$  under Whittle's Index policy.

Replacing  $C^{RP,N}$  by its expression given in Section 4.5 and knowing that  $z_i^{k,*} = \gamma_k u_k^{l_k}(i)$  for  $k \neq m$  and  $z_i^{m,*} = \gamma_m u_m^*(i) = \theta^* \gamma_m u_m^{l_m}(i) + (1 - \theta^*) \gamma_m u_m^{l_m-1}(i)$  (by definition of  $\mathbf{z}^*$ ), then the difference between  $C_T^N(\mathbf{x})$  and  $C^{RP,N}$  can be expressed as:

$$C_T^N(\mathbf{x}) - C^{RP,N} = \left| \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k Z_i^{k,N}(t) i N \mid \mathbf{x} \right] - \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k z_i^{k,*} i N \right] \right| \quad (\text{B.71})$$

We divide all by  $N$

$$\begin{aligned} \frac{C_T^N(\mathbf{x})}{N} - \frac{C^{RP,N}}{N} &= \left| \frac{1}{T} \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L \mathbb{E}(a_k Z_i^{k,N}(t) i) - a_k z_i^{k,*} i \right| \\ &\leq \left| \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^L \mathbb{E}(a_k Z_i^{k,N}(t) i) - a_k z_i^{k,*} i \right| + \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L \mathbb{E}(a_k Z_i^{k,N}(t) i) - a_k z_i^{k,*} i \right| \\ &\leq \frac{T_0 L(L+1)}{T} \sum_{k=1}^K a_k \gamma_k + \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L \mathbb{E}(a_k Z_i^{k,N}(t) i) - a_k z_i^{k,*} i \right| \end{aligned} \quad (\text{B.72})$$

We have that the function  $f : \mathbf{z} \rightarrow \sum_{k=1}^K \sum_{i=0}^L a_k z_i^k i$  is lipchitz and continuous, then for an arbitrary small  $\epsilon$ , there exists  $\mu$  such that if  $\|\mathbf{z} - \mathbf{z}^*\| < \mu$ , then  $|f(\mathbf{z}) - f(\mathbf{z}^*)| < \epsilon$ .

We denote  $Y_N$  the event  $\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu$ , we proceed to bound the second term:

$$\begin{aligned}
 \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L E(a_k Z_i^{k,N}(t)i) - a_k z_i^{k,*} i \right| &\leq P_x(Y_N) \frac{1}{T} \sum_{t=T_0}^{T-1} \mathbb{E} \left[ \left| \sum_{k=1}^K \sum_{i=1}^L (a_k Z_i^{k,N}(t)i) - a_k z_i^{k,*} i \right| | Y_N \right] \\
 &\quad + (1 - P_x(Y_N)) \frac{1}{T} \mathbb{E} \left[ \left| \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L (a_k Z_i^{k,N}(t)i) - a_k z_i^{k,*} i \right| | \overline{Y_N} \right] \\
 &\leq \frac{(T - T_0)L(L + 1)}{T} \sum_{k=1}^K a_k \gamma_k P_x(Y_N) + (1 - P_x(Y_N))\epsilon.
 \end{aligned} \tag{B.73}$$

where the above inequality comes from the fact that  $|a_k Z_i^{k,N}(t)i - a_k z_i^{k,*} i| \leq 2\gamma_k a_k i$ . According to Lemma 5.4, we have  $\lim_{N \rightarrow \infty} P_x(Y_N) = 0$ , then:

$$\lim_{N \rightarrow \infty} \left| \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k Z_i^{k,N}(t)iN \mid \mathbf{x} \right] - \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k z_i^{k,*} iN \right] \right| \leq \frac{T_0 L(L + 1)}{T} \sum_{k=1}^K a_k \gamma_k + \epsilon \tag{B.74}$$

This inequality is true  $\forall \epsilon > 0$ , then:

$$\lim_{N \rightarrow \infty} \left| \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k Z_i^{k,N}(t)iN \mid \mathbf{x} \right] - \frac{1}{T} \mathbb{E} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^L a_k z_i^{k,*} iN \right] \right| \leq \frac{T_0 L(L + 1)}{T} \sum_{k=1}^K a_k \gamma_k \tag{B.75}$$

Finally we have:

$$\lim_{T \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} - \frac{C^{RP,N}}{N} = 0 \tag{B.76}$$

## B.17 Proof of Lemma 4.7

We consider any initial state  $(z^1, z^2, \dots, z^K)$ , and we consider only the following possible event (that arises with strictly positive probability): whatever the transmission decision taken, there is no arrivals for all classes up to time  $T = \frac{1}{\alpha}$  ( $A_k(t) = 0$  from  $t = 0$  up till  $T = \frac{1}{\alpha}$  for all  $1 \leq k \leq K$ ).

To that extent, we show that at time  $T$ , we reach the state  $\mathbf{z}_0$ . For that purpose, we divide the queues into  $\frac{1}{\alpha}$  groups denoted by  $G_1, \dots, G_\alpha$  such that  $G_k$  contains a proportion  $\alpha$  of queues with the highest Whittle indices among all queues of the system excluding those of the groups  $G_1, G_2, \dots, G_{k-1}$  at time  $t = 0$ . Based on this, at time slot  $t = 0$ , the queues in  $G_1$  will be scheduled and will transit to state 0 as the number of arrival packets is equal to 0. According to the expressions given in Proposition 4.3, the Whittle index of state 0 is equal to 0 whatever the value of the class. While according to the same Proposition, the Whittle index of state  $n$  strictly higher than 0 is strictly greater than 0 for any class  $k$ . Therefore, regardless of the class, The Whittle index of state  $n$  strictly higher than 0 is greater than that of 0. Bearing that in mind, at time slot  $t = 1$ , the queues in  $G_2$  at state

different than 0 have the highest Whittle's indices among all system's queues. Therefore, these aforementioned queues will be scheduled, and subsequently, all queues in  $G_2$  will be at state 0. In this way, at time slot  $\frac{1}{\alpha}$ , we get all the queues of the system in state 0. Consequently, at time  $T = \frac{1}{\alpha}$ , we attain the desired state which is  $\mathbf{z}_0$ . That concludes the proof.

## B.18 Proof of Theorem 4.6

$$\lim_{T \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} - \frac{C^{RP,N}}{N} = \sum_{k=1}^K \sum_{i=0}^L a_k \mathbb{E} \left[ Z_i^{k,N}(\infty) \right] i - \sum_{k=1}^K \sum_{i=0}^L a_k z_i^{k,*} i \quad (\text{B.77})$$

We have the function  $f : \mathbf{z} \rightarrow \sum_{k=1}^K \sum_{i=0}^L a_k z_i^k i$  is lipchitz and continuous, then for an arbitrary small  $\epsilon$ , there exists  $\mu$  such that if  $\|\mathbf{z} - \mathbf{z}^*\| < \mu$ , then  $|f(\mathbf{z}) - f(\mathbf{z}^*)| < \epsilon$ .

We denote  $U_N$  the event  $\sup \|\mathbf{Z}^N(\infty) - \mathbf{z}^*\| \geq \mu$ , then :

$$\begin{aligned} \left| \sum_{k=1}^K \sum_{i=0}^L a_k \mathbb{E} \left[ Z_i^{k,N}(\infty) \right] i - \sum_{k=1}^K \sum_{i=0}^L a_k z_i^{k,*} i \right| &\leq P(U_N) \mathbb{E} \left[ \left| \sum_{k=1}^K \sum_{i=0}^L (a_k Z_i^{k,N}(\infty) i) - a_k z_i^{k,*} i \right| \middle| U_N \right] \\ &\quad + (1 - P(U_N)) \mathbb{E} \left[ \left| \sum_{k=1}^K \sum_{i=0}^L (a_k Z_i^{k,N}(\infty) i) - a_k z_i^{k,*} i \right| \middle| \overline{U_N} \right] \\ &\leq L(L+1) \sum_{k=1}^K a_k \gamma_k P(U_N) + (1 - P(U_N)) \epsilon \end{aligned} \quad (\text{B.78})$$

According to Lemma 4.9, we have  $\lim_{N \rightarrow \infty} P(U_N) = 0$ , then:

$$\lim_{N \rightarrow \infty} \left| \sum_{k=1}^K \sum_{i=0}^L a_k \mathbb{E} \left[ Z_i^{k,N}(\infty) \right] i - \sum_{k=1}^K \sum_{i=0}^L a_k z_i^{k,*} i \right| \leq \epsilon \quad (\text{B.79})$$

This is true for any  $\epsilon$ . Finally we have:

$$\lim_{N \rightarrow \infty} \left| \lim_{T \rightarrow \infty} \frac{C_T^N(\mathbf{x})}{N} - \frac{C^{RP,N}}{N} \right| = 0 \quad (\text{B.80})$$

That completes the proof.



# Appendix C

## Appendix: Whittle's index policy for minimizing the Age of Information

### C.1 Proof of Theorem 5.1

**Lemma C.1.**  $V(\cdot)$  is increasing with the age.

*Proof.* The proof is based on the Value iteration algorithm. This algorithm consists of computing by iteration  $V_{t+1}(s)$  for all states  $s$  defined by induction as follows:

$$V_{t+1}(s) = \min_a \{C(s, a) + \sum_{s'} Pr(s' | s, a) V_t(s')\} - \theta \quad (\text{C.1})$$

In fact, regardless of the initial value  $V_0(\cdot)$ ,  $V_t(\cdot)$  will converge to the function  $V(\cdot)$ , the solution of the Bellman equation (5.4). Then, the structural properties of  $V(\cdot)$  are the same as  $V_t(\cdot)$ . In this case, it is sufficient for us to show that, for all  $t$ ,  $V_t(\cdot)$  is increasing with the age in order to conclude the same for  $V(\cdot)$ . To that extent, we prove that  $V_t(\cdot)$  is increasing with  $s$  for all  $t$  by induction. In other words, we prove the following property:

$$V_t(s + 1) > V_t(s), \quad t = 0, 1, \dots \quad (\text{C.2})$$

Without loss of generality, taking  $V_0(s) = 0$  for all  $s$ , then  $V_0(\cdot)$  is clearly increasing with  $s$ .

If  $V_t(\cdot)$  is increasing with  $s$ , i.e. (C.2) holds for  $t$ , we prove that it holds for  $t + 1$ . In this regard, we compare between  $V_{t+1}(s + 1)$  and  $V_{t+1}(s)$ . For that purpose, we define  $V_{t+1}^0(s)$  as the value function at state  $s$  when the action prescribed is a passive one (the user stays idle), and  $V_{t+1}^1(s)$  the value function at state  $s$  when the action prescribed is an active one (the user is granted the channel). Therefore:

- If  $L$  is finite:

- For  $s < L$

$$V_{t+1}^0(s) = s + V_t(s + 1) - \theta \quad (\text{C.3})$$

- For  $s = L$

$$V_{t+1}^0(L) = L + V_t(L) - \theta \quad (\text{C.4})$$

and

– For  $s < L$

$$V_{t+1}^1(s) = s + \lambda + pV_t(1) + (1-p)V_t(s+1) - \theta \quad (\text{C.5})$$

– For  $s = L$

$$V_{t+1}^1(L) = L + \lambda + pV_t(1) + (1-p)V_t(L) - \theta \quad (\text{C.6})$$

- If  $L$  is infinite, we have for all  $s$ :

$$V_{t+1}^0(s) = s + V_t(s+1) - \theta \quad (\text{C.7})$$

and

$$V_{t+1}^1(s) = s + \lambda + pV_t(1) + (1-p)V_t(s+1) - \theta \quad (\text{C.8})$$

We have that:

$$V_{t+1}(s) = \min(V_{t+1}^0(s), V_{t+1}^1(s)) - \theta \quad (\text{C.9})$$

Accordingly, when  $L$  is finite, we have for all  $1 \leq s \leq L-2$ ,  $V_t(s+1) \leq V_t(s+2)$  since  $V_t(\cdot)$  is increasing in  $[1, L]$ . Therefore,  $V_{t+1}^0(\cdot)$  and  $V_{t+1}^1(\cdot)$  are increasing with  $s$  in  $[1, L-1]$ . Furthermore, we have  $V_{t+1}^0(L-1) = V_{t+1}^0(L)$  and  $V_{t+1}^1(L-1) = V_{t+1}^1(L)$ . Consequently,  $V_{t+1}^0(\cdot)$  and  $V_{t+1}^1(\cdot)$  are increasing with  $s$  in  $[1, L]$ . This implies that,  $V_{t+1}(s) = \min(V_{t+1}^0(s), V_{t+1}^1(s))$  is less than  $\min(V_{t+1}^0(s+1), V_{t+1}^1(s+1)) = V_{t+1}(s+1)$ . Hence,  $V_{t+1}(s+1) \geq V_{t+1}(s)$  for  $s \in [1, L-1]$ . Analogously to the case above, when  $L$  is infinite, as  $V_t(\cdot)$  is increasing with  $s$ , then  $V_{t+1}^0(\cdot)$  and  $V_{t+1}^1(\cdot)$  are increasing with  $s$ . Hence,  $V_{t+1}(s+1) \geq V_{t+1}(s)$ . Thus, we have shown by induction that  $V_t(\cdot)$  is increasing with  $s$  for both cases. Consequently, we can deduce that  $V(\cdot)$  will be also increasing with  $s$ . Hence, the proof is complete.  $\square$

In the sequel, we focus only on the case when  $L$  is infinite, since the demonstration for the other case is practically the same. For that, let us define  $\Delta V_{t+1}(s)$  as:

$$\Delta V_{t+1}(s) = V_{t+1}^1(s) - V_{t+1}^0(s) \quad (\text{C.10})$$

Then, by substituting the equations (C.7) and (C.8) in (C.10), we have that:

$$\Delta V_{t+1}(s) = \lambda + pV_t(1) - pV_t(s+1) \quad (\text{C.11})$$

As it was previously discussed,  $\Delta V_{t+1}(s)$  will converge to  $\Delta V(s)$  that equals to  $\lambda + pV(1) - pV(s+1)$ . According to Lemma C.1,  $V(\cdot)$  is increasing with  $s$ , then,  $\Delta V(\cdot)$  is decreasing with  $s$ . This means that there exists a certain state  $s^* \in [0, +\infty]$  such that for all  $s < s^*$ ,  $V^1(s) \geq V^0(s)$  (the optimal action is a passive one) and for all  $s \geq s^*$ ,  $V^1(s) \leq V^0(s)$  (the optimal action is an active one). In other words, there exists a given state  $s^*$  such that for all  $s < s^*$ , remaining idle is more beneficial than transmitting, and for all  $s \geq s^*$ , transmitting is more gainful than staying idle. Consequently, the optimal solution of the Bellman equation (5.4) is an increasing threshold-based policy.

## C.2 Proof of Proposition 5.2

According to Proposition 3.2, we just need to prove that  $\sum_{i=1}^{n-1} u^n(i)$  is increasing with  $n$ . In fact, one can check easily from equations (5.10), (5.12) and (5.14), that  $\sum_{i=n}^L u^n(i)$  is decreasing with  $n$  for both cases of  $L$ . Thereby,  $\sum_{i=1}^{n-1} u^n(i) = 1 - \sum_{i=n}^L u^n(i)$  is increasing with  $n$ . That concludes the proof.

## C.3 Proof of Theorem 5.3

Given that  $\sum_{i=1}^L u^n(i)i$  is strictly increasing with  $n$ , then the Whittle's index of state  $s$  is effectively the one given by the algorithm 1. Moreover  $W_i^k$  in nothing but  $x_{i+1,i}^k$  defined in Lemma 4.3. Subsequently, according to the proof of Theorem 4.3 in Chapter 4, it is sufficient to show that  $W_i^k$  is increasing with  $i$  to establish the desired result.

**Lemma C.2.**  $W_i^k$  is increasing with  $i$  for both cases of  $L$ .

*Proof.* We have these following outcomes:

If  $L < +\infty$ :

$$W_{i+1}^k - W_i^k = (ip_k + 1)(1 - (1 - p_k)^{L-i-1}) \geq 0 \quad (\text{C.12})$$

If  $L = \infty$ :

$$W_{i+1}^k - W_i^k = ip_k + 1 \geq 0 \quad (\text{C.13})$$

That concludes the proof. □

Leveraging the above result,  $W_i^k$  is genuinely the Whittle's index of state  $i$  in the class  $k$ .

## C.4 Proof of Proposition 5.4

Our proof is based on finding the characteristic polynomial of  $Q$ , and investigating the norm of its roots. By examining the expression of  $Q$ , we can deduce that the characteristic polynomial of  $Q$  is the product of the characteristic polynomial of each matrix  $Q_k$ :

$$\chi_Q(\lambda) = \prod_{k=1}^K \chi_{Q_k}(\lambda) \quad (\text{C.14})$$

Therefore, it is sufficient to find the eigenvalues of each matrix  $Q_k$  to conclude those of  $Q$ . To do so, we distinguish between two cases:

$k \neq m$

The characteristic polynomial of the matrix  $Q_k$  is defined as follows:

$$\chi_{Q_k} = \det(Q_k - \lambda I) \quad (\text{C.15})$$

The characteristic polynomial of  $Q_k$  is reported in Table C.2. In order to get a closed-form of this determinant, we apply elementary row and column operations. More specifically, let us denote by  $r_i$  the row  $i$  of the determinant. We also denote by  $a_{i,j}$  the element in row

$$Q_k = \begin{matrix} & 1 & 2 & \cdots & l_k^* - 3 & l_k^* - 2 & l_k^* & l_k^* + 1 & \cdots & L - 1 & L \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ l_k^* - 3 \\ l_k^* - 2 \\ l_k^* \\ l_k^* + 1 \\ \vdots \\ L - 1 \\ L \end{matrix} & \left( \begin{array}{cccccccccccc} 0 & 0 & \cdots & 0 & 0 & p_k & p_k & \cdots & \cdots & p_k \\ 1 & \ddots & & & \vdots & 0 & 0 & & & & 0 \\ 0 & \ddots & \ddots & & & \vdots & \vdots & & & & \vdots \\ 0 & \ddots & 1 & \ddots & & \vdots & \vdots & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & \cdots & \cdots & \cdots & -1 & -1 & \cdots & \cdots & \cdots & \cdots & -1 \\ 0 & 0 & \cdots & \cdots & 0 & 1 - p_k & 0 & & 0 & & 0 \\ \vdots & & & & \vdots & 0 & \ddots & \ddots & & & 0 \\ \vdots & & & & \vdots & & \ddots & \ddots & 0 & & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & 0 & 1 - p_k & 1 - p_k \end{array} \right) \end{matrix}$$
  

$$Q_m = \begin{matrix} & 1 & 2 & \cdots & l_m - 2 & l_m - 1 & l_m + 1 & l_m + 2 & \cdots & L - 1 & L \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ l_m - 2 \\ l_m - 1 \\ l_m + 1 \\ l_m + 2 \\ \vdots \\ L - 1 \\ L \end{matrix} & \left( \begin{array}{cccccccccccc} 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & \ddots & & & \vdots & 0 & 0 & & & & 0 \\ 0 & \ddots & \ddots & & & \vdots & \vdots & & & & \vdots \\ 0 & \ddots & 1 & \ddots & & \vdots & \vdots & & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ -1 & \cdots & \cdots & \cdots & -1 & p_m - 1 & \cdots & \cdots & \cdots & \cdots & p_m - 1 \\ 0 & 0 & \cdots & \cdots & 0 & 1 - p_m & 0 & & 0 & & 0 \\ \vdots & & & & \vdots & 0 & \ddots & \ddots & & & 0 \\ \vdots & & & & \vdots & & \ddots & \ddots & 0 & & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 0 & 0 & 1 - p_m & 1 - p_m \end{array} \right) \end{matrix}$$

Table C.1 – The expressions of the matrices  $Q_k$  for  $k \neq m$  and  $Q_m$

---

$i$  and column  $j$  of the matrix  $Q_k$ . For  $i = 2$  till  $l_k^* - 2$ , we add to each row the previous row multiplied by  $\frac{1}{\lambda}$ . In other words:

$$r_i = \frac{r_{i-1}}{\lambda} + r_i \quad i = 2, \dots, l_k^* - 2 \quad (\text{C.16})$$

Note that these operations are not done simultaneously but rather successively. In other words, in the increasing order of the rows, and at each iteration  $i$ , we add to  $r_i$  the updated row  $r_{i-1}$  at iteration  $i - 1$  multiplied by  $\frac{1}{\lambda}$ . After doing so, we execute the following operation in order to have zeros for the elements  $a_{l_k^*, 1}$  to  $a_{l_k^*, l_k^* - 2}$ :

$$r_{l_k^*} = - \sum_{i=1}^{l_k^* - 2} \frac{r_i}{\lambda} + r_{l_k^*} \quad (\text{C.17})$$

As a result,  $\chi_{Q_k}(\lambda)$  will be the determinant of the matrix  $\mathbf{G}$  reported in the same table. Since  $\mathbf{G}$  is an upper triangular block matrix, we will not be interested in the expression

$$\chi_{Q_k}(\lambda) = \begin{array}{c} 1 \\ 2 \\ \vdots \\ l_k^* - 3 \\ l_k^* - 2 \\ l_k^* \\ l_k^* + 1 \\ \vdots \\ L - 1 \\ L \end{array} \begin{array}{c} 1 \quad 2 \quad \dots \quad l_k^* - 3 \quad l_k^* - 2 \quad l_k^* \quad l_k^* + 1 \quad \dots \quad L - 1 \quad L \\ \left| \begin{array}{cccccccccccc} -\lambda & 0 & \dots & 0 & 0 & p_k & p_k & \dots & \dots & p_k \\ 1 & -\lambda & & & \vdots & 0 & 0 & & & 0 \\ \vdots & 0 & \ddots & \ddots & & \vdots & \vdots & & & \vdots \\ l_k^* - 3 & 0 & \ddots & 1 & \ddots & \vdots & \vdots & & & \vdots \\ l_k^* - 2 & 0 & \dots & 0 & 1 & -\lambda & 0 & 0 & \dots & \dots & 0 \\ l_k^* & -1 & \dots & \dots & \dots & -1 & -1 - \lambda & \dots & \dots & \dots & -1 \\ l_k^* + 1 & 0 & 0 & \dots & \dots & 0 & 1 - p_k & -\lambda & & 0 & 0 \\ \vdots & \vdots & & & & \vdots & 0 & \ddots & \ddots & & 0 \\ L - 1 & \vdots & & & & \vdots & & \ddots & \ddots & -\lambda & 0 \\ L & 0 & \dots & \dots & \dots & 0 & 0 & 0 & 0 & 1 - p_k & 1 - p_k - \lambda \end{array} \right| \end{array}$$

$$\mathbf{G} = \left( \begin{array}{c|cccc} -\lambda & 0 & \dots & 0 & 0 \\ 0 & -\lambda & & & \vdots \\ 0 & \ddots & \ddots & & \vdots \\ 0 & \ddots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 0 & -\lambda \\ \hline \mathbf{0} & -\lambda - 1 - \frac{1}{\lambda} \sum_{i=0}^{l_k^*-2} \frac{p_k}{\lambda^i} & -1 - \frac{1}{\lambda} \sum_{i=0}^{l_k^*-2} \frac{p_k}{\lambda^i} & \dots & \dots & \dots & -1 - \frac{1}{\lambda} \sum_{i=0}^{l_k^*-2} \frac{p_k}{\lambda^i} \\ & 1 - p_k & -\lambda & 0 & \dots & \dots & 0 \\ & 0 & \ddots & \ddots & & & \vdots \\ & \vdots & & \ddots & \ddots & & \vdots \\ & \vdots & & & & -\lambda & 0 \\ & 0 & \dots & \dots & \dots & 1 - p_k & 1 - p_k - \lambda \end{array} \right) \mathbf{B}$$

Table C.2 – The expressions of the characteristic polynomials of  $Q_k$

---

of  $\mathbf{B}$  as the determinant will be independent of it. By letting  $A = 1 + \frac{1}{\lambda} \sum_{i=0}^{l_k^*-2} \frac{p_k}{\lambda^i}$ , the determinant of  $\mathbf{G}$  will be equal to  $(-\lambda)^{l_k^*-2}$  (the determinant of the upper left block) times the following determinant:

$$S_k = \begin{vmatrix} -\lambda - A & -A & \cdots & \cdots & \cdots & -A \\ 1 - p_k & -\lambda & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & & -\lambda & 0 \\ 0 & \cdots & \cdots & \cdots & 1 - p_k & 1 - p_k - \lambda \end{vmatrix}$$

which originates from the lower right block matrix of  $\mathbf{G}$ , and of dimension  $(L - l_k^* + 1) \times (L - l_k^* + 1)$ . Now, we need to find an explicit expression of this determinant, which we will denote by  $S_k$ . To achieve this goal, we first start by developing the determinant through the last column. By doing so, we end up with:

$$S_k = (-1)^{L-l_k^*} (-A)(1 - p_k)^{L-l_k^*} + (-\lambda + 1 - p_k) D_{L-l_k^*} \quad (\text{C.18})$$

where  $D_n$  is determinant of the following matrix:

$$D_n = \begin{vmatrix} & 1 & 2 & \cdots & \cdots & n \\ 1 & \left( \begin{array}{c} -\lambda - A \\ 1 - p_k \\ 0 \\ \vdots \\ \vdots \\ 0 \end{array} \right. & \left( \begin{array}{c} -A \\ -\lambda \\ \ddots \\ \ddots \\ \cdots \end{array} \right) & \left( \begin{array}{c} \cdots \\ 0 \\ \cdots \\ \cdots \\ 1 - p_k \end{array} \right) & \left( \begin{array}{c} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \end{array} \right) & \left( \begin{array}{c} -A \\ 0 \\ \vdots \\ 0 \\ -\lambda \end{array} \right) \end{vmatrix}$$

We need to find the expression of  $D_n$  for any fixed integer value  $n$  in order to conclude it for  $D_{L-l_k^*}$ . To do so, we define  $\Delta_n$  as:

$$\Delta_n = \begin{vmatrix} -\lambda - A & -A & \cdots & \cdots & -A \\ 1 & -\frac{\lambda}{1-p_k} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ & \ddots & \ddots & & 0 \\ 0 & \cdots & \cdots & 1 & -\frac{\lambda}{1-p_k} \end{vmatrix}$$

Hence,  $D_n = (1 - p_k)^{n-1} \Delta_n$ .

The second step consists of computing  $\Delta_n$  for all  $n$ . To that end, we provide the following lemma.

**Lemma C.3.** *The determinant  $\Delta_n$  can be expressed for any integer value  $n$  as:*

$$\Delta_n = (-1)^n \sum_{j=0}^{n-2} \left( \frac{\lambda}{1-p_k} \right)^j A + (-1)^{n+1} \left( \frac{\lambda}{1-p_k} \right)^{n-1} \Delta_1 \quad n \geq 2 \quad (\text{C.19})$$

where  $\Delta_1 = -\lambda - A$ .

*Proof.* The proof follows a mathematical induction. For  $n = 1$ , we can easily check that  $\Delta_1$  is indeed  $-\lambda - A$  by replacing  $n$  by 1. Next, we suppose that (C.19) holds up till  $n$  and we aim to prove that it holds for  $n + 1$ . By developing  $\Delta_{n+1}$  through the last column, we get:

$$\Delta_{n+1} = (-1)^{n+2}(-A) - \frac{\lambda}{1-p_k}\Delta_n \quad (\text{C.20})$$

By replacing  $\Delta_n$  with its value, we end up with:

$$\begin{aligned} \Delta_{n+1} &= (-1)^{n+2}(-A) - \frac{\lambda}{1-p_k} \left[ (-1)^n \sum_{j=0}^{n-2} \left( \frac{\lambda}{1-p_k} \right)^j A + (-1)^{n+1} \left( \frac{\lambda}{1-p_k} \right)^{n-1} \Delta_1 \right] \\ &= (-1)^{n+2}(-A) + (-1)^{n+1} \sum_{j=0}^{n-2} \left( \frac{\lambda}{1-p_k} \right)^{j+1} A + (-1)^n \left( \frac{\lambda}{1-p_k} \right)^n \Delta_1 \\ &= (-1)^{n+2}(-A) + (-1)^{n+1} \sum_{j=1}^{n-1} \left( \frac{\lambda}{1-p_k} \right)^j A + (-1)^n \left( \frac{\lambda}{1-p_k} \right)^n \Delta_1 \\ &= (-1)^{n+1} A + (-1)^{n+1} \sum_{j=1}^{n-1} \left( \frac{\lambda}{1-p_k} \right)^j A + (-1)^n \left( \frac{\lambda}{1-p_k} \right)^n \Delta_1 \\ &= (-1)^{n+1} \sum_{j=0}^{n-1} \left( \frac{\lambda}{1-p_k} \right)^j A + (-1)^{n+2} \left( \frac{\lambda}{1-p_k} \right)^n \Delta_1 \end{aligned}$$

Therefore, (C.19) holds for  $n + 1$  which concludes our proof. □

By leveraging the above lemma, we can conclude the expression of  $D_{L-l_k^*}$ :

$$D_{L-l_k^*} = (1-p_k)^{L-l_k^*-1} \left[ (-1)^{L-l_k^*} \sum_{j=0}^{L-l_k^*-2} \left( \frac{\lambda}{1-p_k} \right)^j A + (-1)^{L-l_k^*+1} \left( \frac{\lambda}{1-p_k} \right)^{L-l_k^*-1} \Delta_1 \right] \quad (\text{C.21})$$

As a consequence, we are able to find  $S_k$  in function of  $\lambda$  and  $p_k$ , as well as the expression of  $\chi_{Q_k}(\lambda)$ . In fact, after computations, we get:

$$\chi_{Q_k}(\lambda) = (-1)^{L+1} \lambda^{L-l_k^*} \left( \lambda^{l_k^*-1} + \sum_{i=0}^{l_k^*-2} p_k \lambda^i \right) \quad (\text{C.22})$$

Now, based on the expression of the characteristic polynomial  $\chi_{Q_k}$ , we prove that all the eigenvalues of  $Q_k$  have a modulus strictly less than one. We prove this result by contradiction. More specifically, we suppose there exists a given eigenvalue of the matrix  $Q_k$  that satisfies  $|\lambda| \geq 1$ . As  $\lambda$  is an eigenvalue of  $Q_k$ , it is therefore a root of  $\chi_{Q_k}(\lambda)$ . Hence, it verifies:

$$\lambda^{l_k^*-1} = - \sum_{i=0}^{l_k^*-2} p_k \lambda^i = -p_k \frac{1 - \lambda^{l_k^*-1}}{1 - \lambda} \quad (\text{C.23})$$

By factorizing the element  $\lambda^{l_k^*}$ , and by using the modulus on both sides, we get:

$$\begin{aligned} p_k &= |\lambda|^{l_k^*-1} |p_k - 1 + \lambda| \\ &\stackrel{(a)}{\geq} |\lambda|^{l_k^*-1} (|\lambda| - |1 - p_k|) \\ &\stackrel{(b)}{\geq} |\lambda| (|\lambda| - |1 - p_k|) \\ &= |\lambda|^2 - |\lambda|(1 - p_k) \end{aligned}$$

where (a) and (b) originate from the reverse triangular inequality and the fact that  $|\lambda| \geq 1$  respectively. Hence:

$$|\lambda|^2 - |\lambda|(1 - p_k) - p_k \leq 0 \quad (\text{C.24})$$

By employing standard real functions analysis, it can be shown that the polynomial:

$$x^2 - (1 - p_k)x - p_k \quad (\text{C.25})$$

is negative if and only if  $x \in [-p_k, 1]$ . However,  $|\lambda| \geq 1$  by assumption. Accordingly,  $|\lambda|$  can only be equal to 1. Next, we prove that, in this case, the imaginary part of  $\lambda$  is equal to zero. To that end, let us consider  $\lambda = x + iy$ . Therefore, we have:

$$p_k = |\lambda|^{l_k^*-1} |p_k - 1 + x + iy| = |p_k - 1 + x + iy| \quad (\text{C.26})$$

By using the definition of the modulus, and by squaring both sides, we get:

$$p_k^2 = (1 - x - p_k)^2 + y^2 \quad (\text{C.27})$$

Knowing that  $x^2 + y^2 = 1$ , we can deduce:

$$\frac{2 - 2p_k}{2(1 - p_k)} = x \quad (\text{C.28})$$

Hence,  $x = 1$ , i.e.  $y = 0$ , and we can deduce that  $\lambda = 1$ . However, 1 is not eigenvalue of matrix  $Q_k$ . This can be seen by replacing  $\lambda$  with 1 in the characteristic polynomial of  $Q_k$ . Accordingly, the hypothesis that  $|\lambda| \geq 1$  fails and all the eigenvalues of  $Q_k$  for any  $k \neq m$  have a modulus strictly less than 1.

$k = m$

The characteristic polynomial of the matrix  $Q_m$  is reported in Table C.3. We follow the same steps of the previous case. For  $i = 2$  till  $l_m^* - 1$ , we sequentially add to each row, the previous row multiplied by  $\frac{1}{\lambda}$ . In other words:

$$r_i = \frac{r_{i-1}}{\lambda} + r_i \quad i = 2, \dots, l_m^* - 1 \quad (\text{C.29})$$

Then, we execute the following operation in order to have zeros for the elements  $a_{l_m^*+1,1}$  to  $a_{l_m^*+1, l_m^*-1}$ :

$$r_{l_m^*+1} = - \sum_{i=1}^{l_m^*-1} \frac{r_i}{\lambda} + r_{l_m^*+1} \quad (\text{C.30})$$



of the value of  $A$ . Given that, we can find the expression of  $S_m$  in function of  $\lambda$  and  $p_m$  as well as the expression of  $\chi_{Q_m}(\lambda)$ . After extensive computations, we obtain:

$$\chi_{Q_m}(\lambda) = (-\lambda)^{L-1} \quad (\text{C.31})$$

Hence, the only eigenvalue of  $Q_m$  is 0 that has a multiplicity of  $L - 1$ .

By combining all these results, we can conclude that the eigenvalues of the matrix  $\mathbf{Q}$  have a modulus strictly less than one, which concludes our proof.

## C.5 Proof of Lemma 5.1

We can formulate the fluid limit equation (5.33) as follows:

$$\mathbf{z}(t+1) = \mathbb{E} \left[ \mathbf{Z}^N(t+1) \middle| \mathbf{Z}^N(t) = \mathbf{z}(t) \right]$$

At time  $t+1$ , applying Whittle index policy, in average exactly a proportion of  $p_k \alpha_k(t)$  of users will be at state one since  $\alpha_k(t)$  refers to the proportion of users in class  $k$  that are scheduled. Accordingly,  $z_1^k(t+1) = p_k \alpha_k(t)$ . While for  $1 \leq i < l_k(t)$ , the users' proportion  $z_i^k(t)$  is not scheduled. Therefore at time  $t+1$ , since prescribing idle action to a given user implies that its state will be increased by 1, the proportion  $z_i^k(t)$  at state  $i$  in class  $k$  will be at state  $i+1$ . Thus,  $\mathbb{E} \left[ Z_{i+1}^{N,k}(t+1) \middle| \mathbf{Z}^N(t) = \mathbf{z}(t) \right] = z_{i+1}^k(t+1) = z_i^k(t)$ .

## C.6 Proof of Lemma 5.2

First of all, we provide an useful lemma.

**Lemma C.4.** *We have for all integer  $i$  and for  $k = 1, 2$ :*

$$w_k(i+1) - w_k(i) = ip_k + 1$$

*Proof.* The result can be obtained directly by replacing  $w_k(i)$  by its expression. □

In order to prove the present lemma, we proceed in two steps:

- We prove first by contradiction that there exists a given time  $t_f$  such that  $\alpha_1(t_f) > 0$ .
- We prove that if  $\alpha_1(t_f) > 0$ , then  $\alpha_1(t) > 0$  for all  $t \geq t_f$ .

1. For the first point, we suppose that for all  $t$ , we have that  $\alpha_1(t) = 0$ . Consequently, we get that  $z_1^1(t+T_t) = 0, \dots, z_{l_1(t+T_t)}^1(t+T_t) = 0$ , and  $\alpha_1(t+T_t) = 0$ . This means that, the proportion of all users in class 1 is equal to 0. However, the users' proportion of class 1 is  $\gamma_1 \neq 0$ . That is, there exists a given time  $t_f$  such  $\alpha_1(t_f) > 0$ .

2. Before addressing the second point, we recall that  $\alpha_1(t)$  refers to the scheduled users' proportion in the class 1. Thereby,  $\alpha_1(t)$  contains all users with the highest Whittle index values among all users in class 1. To that extent, at time  $t_f$ , the Whittle index of  $\alpha_1(t_f)$  is greater than the Whittle index of the users' proportion  $1 - \alpha$  that we

denote by  $C$ . We let  $\mathcal{S}_{t_f}(C)$  be the set of pair (state,class) at time  $t_f$  in the users' proportion  $C$ . Denoting by  $q$  the smallest state of  $\alpha_1(t_f)$ ,  $n$  and  $m$  a given state and class respectively such that  $z_n^m(t)$  belongs to  $C$  at time  $t_f$ , then  $w_m(n) \leq w_1(q)$ . Under the Whittle index policy, at time  $t_f + 1$ , the states of a users' proportion that equals to  $(1 - p_1)\alpha_1(t_f)$  among the users' proportion  $\alpha_1(t_f)$ , will be increased by one in comparison with the time slot  $t_f$ , as well as the users' proportion  $C$ . Accordingly, the smallest state of the proportion  $(1 - p_1)\alpha_1(t_f)$ , is  $q + 1$ .  $\mathcal{S}_{t_f+1}(C)$  is shifted of one with respect to  $\mathcal{S}_{t_f}(C)$ , i.e.,  $(n, m) \in \mathcal{S}_{t_f}(C) \Leftrightarrow (n + 1, m) \in \mathcal{S}_{t_f+1}(C)$ . We compare  $w_1(q + 1)$  with the Whittle index of  $n$  in class  $m$  such that  $(n, m) \in \mathcal{S}_{t_f+1}(C)$ . In that direction, we let  $(n, m) \in \mathcal{S}_{t_f+1}(C)$ , and we distinguish between two cases:

- $m = 1$ : Leveraging the fact that  $(n - 1, m) \in \mathcal{S}_{t_f}(C)$ , then  $w_1(q) \geq w_1(n - 1)$ . That implies that  $n - 1 \leq q$  since  $w_k(\cdot)$  is increasing. Hence  $n \leq q + 1$ . As consequence,  $w_1(n) \leq w_1(q + 1)$
- $m = 2$ : Again we distinguish between two case:
  - If  $n - 1 \leq q$ , then  $w_2(n) < w_1(n) \leq w_1(q + 1)$ .  
Therefore, we obtain our desired result for the first case.
  - If  $n - 1 > q$ :  
We have that:

$$w_1(q+1) - w_2(n) = (w_1(q+1) - w_1(q)) - (w_2(n) - w_2(n-1)) + w_1(q) - w_2(n-1)$$

Applying Lemma C.4, we obtain:  $(w_1(q+1) - w_1(q)) - (w_2(n) - w_2(n-1)) = qp_1 - (n - 1)p_2$ . Given that  $w_2(n - 1) \leq w_1(q)$ , therefore replacing by their expressions we get:

$$(n - 2)(n - 1)p_2/2 + n - 1 \leq (q - 1)qp_1/2 + q$$

As  $n - 1 > q$ , then:

$$(n - 2)(n - 1)p_2/2 \leq (q - 1)qp_1/2$$

Hence:

$$(n - 1)p_2 \leq qp_1$$

Therefore,  $(w_1(q + 1) - w_1(q)) - (w_2(n) - w_2(n - 1)) \geq 0$ . Hence, knowing that  $w_1(q) - w_2(n - 1) \geq 0$  we end up with our desired result for this case, i.e.  $w_1(q + 1) - w_2(n) \geq 0$ .

Thus, we have proved that at time  $t_f + 1$ , all the users' proportions in  $C$  whose sum is equal to  $1 - \alpha$  have a Whittle index less than that of  $(1 - p_1)\alpha_1(t_f)$  defined in the beginning of this proof. That means that there exists at least a users' proportion that equals to  $1 - \alpha$  with Whittle index values less than those of the states of the users' proportion  $(1 - p_1)\alpha_1(t_f)$ . Then surely, the users' proportion  $(1 - p_1)\alpha_1(t_f)$  that is different from 0 belongs to the users' proportion  $\alpha$  with the highest Whittle index values. This implies that surely at time  $t_f + 1$ , there will be at least one queue in class 1 belonging to  $\alpha$  with the highest Whittle index values. Therefore, we have that  $\alpha_1(t_f + 1) > 0$ . This result can be generalized for all  $t \geq t_f$ . In other words, we have for all  $t \geq t_f$ ,  $\alpha_1(t) > 0$ .

## C.7 Proof of Lemma 5.3

As  $\alpha_1(j) + \alpha_2(j) = \alpha$  for all integers  $j$ , then, if  $\alpha_1(t+i) = 0$ ,  $\alpha_2(t+i) = \alpha$ . For  $j \in [1, T_{\max}]$ , we have that  $T_{\max} - j \in [0, T_{\max} - 1]$ . This means that  $\alpha_1(t + T_{\max} - j)$  is equal to 0, which implies that  $\alpha_2(t + T_{\max} - j) = \alpha$ . Moreover, knowing that  $l(t + T_{\max}) \leq T_{\max}$ , then for all  $j \in [1, l(t + T_{\max})]$ ,  $T_{\max} - j \in [T_{\max} - l(t + T_{\max}), T_{\max} - 1] \subset [0, T_{\max} - 1]$ . Hence, we get that  $\alpha_1(t + T_{\max} - j) = 0$ , for all  $j \in [1, l(t + T_{\max})]$ .

Therefore, according to the definition 5.2,  $T_{\max}$  satisfies:

$$T_{\max} p_2 \alpha \geq 1 - \alpha \quad (\text{C.32})$$

$$T_{\max} \geq \frac{1 - \alpha}{p_2 \alpha} \quad (\text{C.33})$$

Providing that  $T_{\max}$  by definition is the first time when  $\sum_{j=1}^{l(t+T_{\max})} p_1 \alpha_1(t + T_{\max} - j) + \sum_{j=1}^{T_{\max}} p_2 \alpha_2(t + T_{\max} - j)$  exceeds  $1 - \alpha$ , then at time  $t + T_{\max} - 1$ ,  $\sum_{j=1}^{l(t+T_{\max}-1)} p_1 \alpha_1(t + T_{\max} - 1 - j) + \sum_{j=1}^{T_{\max}-1} p_2 \alpha_2(t + T_{\max} - 1 - j) < 1 - \alpha$ . This latter sum is equal to  $(T_{\max} - 1) p_2 \alpha$  which is less than  $1 - \alpha$ . Therefore, we have as result that  $T_{\max} < \frac{1-\alpha}{p_2 \alpha} + 1$ . As there is one integer value between  $\frac{1-\alpha}{p_2 \alpha}$  and  $\frac{1-\alpha}{p_2 \alpha} + 1$ , then  $T_{\max}$  doesn't depend on  $t$ , and satisfies:  $\frac{1-\alpha}{p_2 \alpha} \leq T_{\max} < \frac{1-\alpha}{p_2 \alpha} + 1$ .

## C.8 Proof of Proposition 5.8

We have that  $w_1(n) = \frac{(n-1)p_1 n}{2} + n$ , and  $w_2(n) = \frac{(n-1)p_2 n}{2} + n$ . We start first by finding the set of states for which the Whittle index alternate between the two classes. As we can see from the expression of the Whittle index, for a given state  $n$ ,  $w_2(n) < w_1(n)$  as  $p_2 < p_1$ . In order to have the condition of alternation strictly satisfied for any given state  $n$ , we must have  $w_1(n) < w_2(n+1)$ . Hence, denoting by  $f(n)$  the difference  $w_2(n+1) - w_1(n)$ , we study the sign of  $f(n)$  to see for which  $n$   $f$  is strictly positive.

**Lemma C.5.** *For all  $n \in [1, D[$ ,  $f(n) > 0$*

*Proof.* We have that:

$$f(n) = \frac{n^2}{2}(p_2 - p_1) + \frac{n}{2}(p_1 + p_2) + 1 \quad (\text{C.34})$$

Hence:

$$f'(n) = n(p_2 - p_1) + \frac{p_1 + p_2}{2} \quad (\text{C.35})$$

The derivative is equal to zero for  $n = \frac{p_1 + p_2}{2(p_1 - p_2)}$ , which is strictly greater than 0. This means that  $f$  is strictly increasing in  $[0, \frac{p_1 + p_2}{2(p_1 - p_2)}]$  since  $f'(n) > 0$  in  $[0, \frac{p_1 + p_2}{2(p_1 - p_2)}[$ . Providing that  $f(0) = 1$ , then surely  $f$  is strictly positive in  $[0, \frac{p_1 + p_2}{2(p_1 - p_2)}]$ . This means that, the unique positive solution for  $f(n) = 0$  must be in the interval  $[\frac{p_1 + p_2}{2(p_1 - p_2)}, +\infty[$ , as  $\lim_{n \rightarrow +\infty} f(n) = -\infty$ .

Indeed, the unique solution  $n_0$  of  $f(n) = 0$  in  $[\frac{p_1 + p_2}{2(p_1 - p_2)}, +\infty[$  is the biggest root of the polynomial (C.34) which is exactly the value  $D$  introduced in Assumption 5.1. As the function  $f$  is decreasing in  $[\frac{p_1 + p_2}{2(p_1 - p_2)}, +\infty[$ , then  $f$  is strictly positive in  $[0, D[$ . Therefore,  $f(n) > 0$  for  $n \in [1, D[$ , which concludes the proof. ■

According to Lemma C.5, the order of the Whittle index strictly alternates between the two states when  $n \in [1, D[$ . Therefore, we need to prove that  $T_{\max} + 1$  is upper bounded by  $D$  in order to prove that the alternation condition is satisfied from state 1 to  $T_{\max} + 1$ . Indeed, as we have found an upper bound of  $T_{\max}$  which is equal to  $\frac{1-\alpha}{p_2\alpha} + 1$  (according to Lemma 5.3), we just need to prove that  $\frac{1-\alpha}{p_2\alpha} + 2$  is strictly less than  $D$ .

Under assumption (5.1), we have that:

$$\alpha > \frac{1}{1 + (D - 2)p_2} \quad (\text{C.36})$$

$$\alpha(1 + p_2(D - 2)) > 1 \quad (\text{C.37})$$

$$\alpha p_2(D - 2) > 1 - \alpha \quad (\text{C.38})$$

$$D - 2 > \frac{1 - \alpha}{p_2\alpha} \quad (\text{C.39})$$

$$D > \frac{1 - \alpha}{p_2\alpha} + 2 \quad (\text{C.40})$$

Hence, from state 1 to  $T_{\max} + 1$ , the order of the Whittle index strictly alternates between the two classes. Accordingly, the proof is concluded.

## C.9 Proof of Proposition 5.9

We present first a lemma which will be helpful in proving this proposition as well as the next ones.

**Lemma C.6.** *For any state  $q$ , at any time  $t$ , we have that:*

$$w_1(q) \leq w_2(l_2(t)) \Rightarrow w_1(q) \leq w_1(l_1(t))$$

and

$$w_2(q) \leq w_1(l_1(t)) \Rightarrow w_2(q) \leq w_2(l_2(t))$$

*Proof.* See appendix C.10 ■

We consider  $t \geq t_f$ . After time  $T_t$ , we have that:

$$\sum_{j=1}^{l(t+T_t)} p_1\alpha_1(t + T_t - j) + \sum_{j=1}^{T_t} p_2\alpha_2(t + T_t - j) \geq 1 - \alpha \quad (\text{C.41})$$

Then, as it has been showcased, at time  $t + T_t$ , there exists  $l_1(t + T_t) \leq l(t + T_t)$ ,  $l_2(t + T_t) \leq T_t$ ,  $\gamma(t + T_t) = 1$  and  $0 < \beta(t + T_t) \leq 1$ ; or  $0 < \gamma(t + T_t) \leq 1$  and  $\beta(t + T_t) = 1$  such that:

$$\begin{aligned} & \sum_{j=1}^{l_1(t+T_t)-1} p_1\alpha_1(t + T_t - j) + \sum_{j=1}^{l_2(t+T_t)-1} p_2\alpha_2(t + T_t - j) \\ & + \beta(t + T_t)p_1\alpha_1(t + T_t - l_1(t + T_t)) + \gamma(t + T_t)p_2\alpha_2(t + T_t - l_2(t + T_t)) = 1 - \alpha \end{aligned} \quad (\text{C.42})$$

with  $l_1(t+T_t)$  and  $l_2(t+T_t)$  being the instantaneous thresholds in class 1 and 2 respectively at time  $t+T_t$ .

Now, we prove by contradiction that  $\max(l_1(t+T_t), l_2(t+T_t)) \leq T_{max}$ .

We prove first that  $l_2(t+T_t)$  is greater than  $l_1(t+T_t)$ .

As we have that  $w_2(l_1(t+T_t)) < w_1(l_1(t+T_t))$ , then according to lemma C.6,  $w_2(l_1(t+T_t)) \leq w_2(l_2(t+T_t))$ . This implies that  $l_2(t+T_t)$  is greater than  $l_1(t+T_t)$ .

Reasoning by contradiction, we suppose that  $l_2(t+T_t) > T_{max}$  ( $l_2(t+T_t) = \max(l_1(t+T_t), l_2(t+T_t)) > T_{max}$ ). Based on this, we have that  $w_1(T_{max}) < w_2(l_2(t+T_t))$  because  $w_1(T_{max}) < w_2(T_{max} + 1) \leq w_2(l_2(t+T_t))$  since the order of the Whittle index alternates between the two classes as it has been proved in Proposition 5.9. To that extent, we distinguish between two cases:

1) **First case:** If  $\beta(t+T_t) = 1$ :

We have that  $w_1(T_{max}) < w_2(l_2(t+T_t))$ . Then, according to Lemma C.6, we have that  $w_1(T_{max}) \leq w_1(l_1(t+T_t))$ . Hence, we can conclude that  $T_{max} \leq l_1(t+T_t)$  as  $w_1$  is an increasing function with the age of information.

Moreover, since we have that  $p_1\alpha_1(t+T_t-j) + p_2\alpha_2(t+T_t-j) > p_2\alpha$  (the strict inequality is due to the fact that  $\alpha_1(t) > 0$  as  $t \geq t_f$  according to Lemma 5.2), then according to Lemma 5.3, we obtain:

$$\begin{aligned}
 & \sum_{j=1}^{l_1(t+T_t)-1} p_1\alpha_1(t+T_t-j) + \sum_{j=1}^{l_2(t+T_t)-1} p_2\alpha_2(t+T_t-j) \\
 & \quad + \beta(t+T_t)p_1\alpha_1(t+T_t-l_1(t+T_t)) + \gamma(t+T_t)p_2\alpha_2(t+T_t-l_2(t+T_t)) = 1 - \alpha \\
 & = \sum_{j=1}^{l_1(t+T_t)} p_1\alpha_1(t+T_t-j) + \sum_{j=1}^{l_2(t+T_t)-1} p_2\alpha_2(t+T_t-j) + \gamma(t+T_t)p_2\alpha_2(t+T_t-l_2(t+T_t)) \\
 & \geq \sum_{j=1}^{T_{max}} p_1\alpha_1(t+T_t-j) + \sum_{j=1}^{T_{max}} p_2\alpha_2(t+T_t-j) > T_{max}p_2\alpha \geq 1 - \alpha \tag{C.43}
 \end{aligned}$$

The last inequality comes from the fact that  $T_{max} \geq \frac{1-\alpha}{p_2\alpha}$ . This implies that:

$$1 - \alpha > 1 - \alpha \tag{C.44}$$

This gives us an illogical statement. Consequently, in this case, the assumption  $l_2(t+T_t) > T_{max}$  is not true.

2) **Second case:** If  $\beta(t+T_t) < 1$ :

As we have that  $\beta(t+T_t) < 1$ , then  $\gamma(t+T_t)$  should be equal to 1. Therefore, all users at state  $l_2(t+T_t)$  in class 2 are in the users' proportion  $1 - \alpha$  with the smallest Whittle index values. However, there exists users in state  $l_1(t+T_t)$  in class 1 in the users' proportion  $\alpha$  that has the highest Whittle index values. That is, we have  $w_1(l_1(t+T_t)) \geq w_2(l_2(t+T_t))$ . As it has been established before tackling the first case,  $w_1(T_{max}) < w_2(l_2(t+T_t))$ , then

$w_1(T_{\max}) < w_1(l_1(t + T_t))$ . This means that  $l_1(t + T_t) > T_{\max}$ . Therefore, we have that:

$$\begin{aligned} & \sum_{j=1}^{l_1(t+T_t)-1} p_1 \alpha_1(t + T_t - j) + \sum_{j=1}^{l_2(t+T_t)-1} p_2 \alpha_2(t + T_t - j) \\ & + \beta(t + T_t) p_1 \alpha_1(t + T_t - l_1(t + T_t)) + \gamma(t + T_t) p_2 \alpha_2(t + T_t - l_2(t + T_t)) = 1 - \alpha \\ & \geq \sum_{j=1}^{T_{\max}} p_1 \alpha_1(t + T_t - j) + \sum_{j=1}^{T_{\max}} p_2 \alpha_2(t + T_t - j) > T_{\max} p_2 \alpha \geq 1 - \alpha \end{aligned} \quad (\text{C.45})$$

This implies that:

$$1 - \alpha > 1 - \alpha \quad (\text{C.46})$$

Consequently, in this case, the assumption  $l_2(t + T_t) > T_{\max}$  is not true.

Hence, in both cases,  $l_2(t + T_t)$  must be less than  $T_{\max}$ , i.e.  $\max(l_1(t + T_t), l_2(t + T_t)) \leq T_{\max}$  for all  $t$ .

Thus, we end up with  $T_{\max} = l_{\max}$ , which concludes our proof.

## C.10 Proof of Lemma C.6

We prove only the first statement as the proof steps for both cases are exactly the same. By definition of  $l_1(t)$  and  $l_2(t)$ , we have that  $\{z_i^1(t)\}_{1 \leq i \leq l_1(t)} \cup \{z_i^2(t)\}_{1 \leq i \leq l_2(t)}$  is exactly the set  $\{z_i^k(t) : w_k(i) \leq \max(w_1(l_1(t)), w_2(l_2(t)))\}$ . Hence, if a given  $q$  verifies  $w_1(q) \leq w_2(l_2(t))$ , then  $w_1(q) \leq \max(w_1(l_1(t)), w_2(l_2(t)))$ , that implies that  $z_q^1(t) \in \{z_i^k(t) : w_k(i) \leq \max(w_1(l_1(t)), w_2(l_2(t)))\} = \{z_i^1(t)\}_{1 \leq i \leq l_1(t)} \cup \{z_i^2(t)\}_{1 \leq i \leq l_2(t)}$ . Knowing that the highest users' proportion's state of the aforementioned set in class 1 is  $l_1(t)$ , then  $q \leq l_1(t)$ . Therefore as  $w_1(\cdot)$  is increasing,  $w_1(q) \leq w_1(l_1(t))$ .

## C.11 Proof of Proposition 5.10

We have that:

$$\sum_{i=1}^{l_1(t)-1} z_i^1(t) + \sum_{i=1}^{l_2(t)-1} z_i^2(t) + \beta(t) z_{l_1(t)}^1(t) + \gamma(t) z_{l_2(t)}^2(t) = 1 - \alpha \quad (\text{C.47})$$

with  $l_1(t)$  and  $l_2(t)$  being the thresholds in class 1 and 2 respectively at time  $t$ , and  $\beta(t) = 1$  and  $0 < \gamma(t) \leq 1$ , or  $\gamma(t) = 1$  and  $0 < \beta(t) \leq 1$ .

Our aim in this proof is to show that there is a link between  $l_1(t)$  and  $l_2(t)$  when they are less than  $T_{\max}$ . By doing so, we find a general form of the aforementioned equation. To that end, we prove first that  $l_1(t)$  is less than  $l_2(t)$ .

Indeed, as we have  $w_2(l_1(t)) < w_1(l_1(t))$ , then according to lemma C.6,  $w_2(l_1(t)) \leq w_2(l_2(t))$ . Consequently, we can conclude that  $l_1(t) \leq l_2(t)$ .

Secondly, we prove that  $l_2(t) \leq l_1(t) + 1$ . As the order of the Whittle indices alternates between the two classes from state 1 to state  $T_{\max} + 1$ ,  $w_1(l_2(t) - 1) < w_2(l_2(t))$ . Hence, according to lemma C.6, we have that  $w_1(l_2(t) - 1) \leq w_1(l_1(t))$ . Consequently,  $l_2(t) - 1 \leq l_1(t)$ .

Given that  $l_1(t) \leq l_2(t) \leq l_1(t) + 1$ , then  $l_1(t)$  can be either  $l_2(t)$  or  $l_2(t) - 1$ .

The second step consists of deriving the value of  $\beta(t)$  or  $\gamma(t)$  depending on the value of  $l_1(t)$  and  $l_2(t)$ .

- If  $l_1(t) = l_2(t)$ :

We prove that  $\gamma(t) = 1$  if  $z_{l_2(t)}^2 > 0$ . Indeed, if  $\gamma(t) \neq 1$  and  $z_{l_2(t)}^2 > 0$ , thus there is at least a non empty set of users in class 2 at state  $l_2(t)$  that belongs to the users' proportion  $\alpha$  with the highest Whittle index values. However there exists always a non empty set of queues in class 1 at state  $l_1(t)$  that belong to  $1 - \alpha$  users' proportion with the least Whittle index values, since  $\beta(t) > 0$ . Then, we have that  $w_2(l_2(t)) \geq w_1(l_1(t))$ . However, we know that  $w_2(l_2(t)) = w_2(l_1(t)) < w_1(l_1(t))$ . This later inequality contradicts with what precedes. Thus, the statement that  $\gamma(t) \neq 1$  is not true, i.e.  $\gamma(t) = 1$ .

In this case we denote  $l(t) = l_1(t) = l_2(t)$ .

We end up:

$$\sum_{j=1}^{l(t)-1} z_j^1(t) + \sum_{i=1}^{l(t)-1} z_i^2(t) + \beta(t)z_{l(t)}^1(t) + z_{l(t)}^2(t) = 1 - \alpha \quad (\text{C.48})$$

If  $z_{l_2(t)}^2 = 0$ , the last equation still valid since  $z_{l_2(t)}^2 = 0$  whatever the value of  $\gamma(t)$ , namely when  $\gamma(t) = 1$ .

- If  $l_1(t) + 1 = l_2(t)$ :

We prove that  $\beta(t) = 1$  if  $z_{l_1(t)}^1 > 0$ . Indeed, if  $\beta(t) \neq 1$  and  $z_{l_1(t)}^1 > 0$ , there is at least a set of users in class 1 in state  $l_1(t)$  that belongs to the users' proportion  $\alpha$  with the highest Whittle index values. However there is always a set of queues in class 2 at state  $l_2(t)$  that belong to  $1 - \alpha$  users' proportion with the least Whittle index values, since  $\gamma(t) > 0$ . Then, we have that  $w_1(l_1(t)) \geq w_2(l_2(t))$ . However, we know that  $w_2(l_2(t)) = w_2(l_1(t) + 1) > w_1(l_1(t))$  since the order of Whittle index alternates between the two classes from state 1 to  $T_{\max} + 1$  according to Proposition 5.9. Thus,  $w_2(l_1(t) + 1) > w_1(l_1(t)) \geq w_2(l_1(t) + 1)$ , which gives us an obvious contradiction. Therefore, we can assert that  $\beta(t) = 1$ .

In this case, we consider that  $l(t) = l_1(t) + 1 = l_2(t)$  and we get:

$$\sum_{i=1}^{l(t)-1} z_i^1(t) + \sum_{i=1}^{l(t)-1} z_i^2(t) + \gamma(t)z_{l(t)}^2(t) = 1 - \alpha \quad (\text{C.49})$$

Similarly to the first case, if  $z_{l_1(t)}^1 = 0$ , the last equation still valid since  $z_{l_1(t)}^1 = 0$  whatever the value of  $\beta(t)$ , namely when  $\beta(t) = 1$ . Subsequently, combining the two cases, there exists  $l(t)$  such that:

$$\sum_{i=1}^{l(t)-1} z_i^1(t) + \sum_{i=1}^{l(t)-1} z_i^2(t) + \beta(t)z_{l(t)}^1(t) + \gamma(t)z_{l(t)}^2(t) = 1 - \alpha \quad (\text{C.50})$$

where  $\beta(t) = 0$  and  $0 < \gamma(t) \leq 1$ , or  $0 < \beta(t) \leq 1$  and  $\gamma(t) = 1$

## C.12 Proof of Proposition 5.11

We prove the Proposition by induction:

- For  $T = T_0$ , we have already proved our claim.
- We suppose that the statement is valid for a given  $T$ , i.e. there exists  $l(T)$ ,  $\beta(T)$  and  $\gamma(T)$  such that:

$$\sum_{j=1}^{l(T)-1} p_1 \alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2 \alpha_2(T-j) + \beta(T) p_1 \alpha_1(T-l(T)) + \gamma(T) p_2 \alpha_2(T-l(T)) = 1 - \alpha \tag{C.51}$$

where  $\beta(T) = 0$  and  $0 < \gamma(T) \leq 1$ , or  $0 < \beta(T) \leq 1$  and  $\gamma(T) = 1$ . Then, at the next time slot, among the users' proportion scheduled,  $\alpha$ , exactly  $p_1 \alpha_1(T)$  and  $p_2 \alpha_2(T)$  will go to state one for each class, while for the rest, their states will be incremented by one. Likewise, for the other users for which the action taken is passive, their states will be incremented by one. As consequence, the decreasing order according to the Whittle index value for these proportions of users at the next slot is  $\beta(T) p_1 \alpha_1(T-l(T))$ ,  $\gamma(T) p_2 \alpha_2(T-l(T))$ ,  $p_1 \alpha_1(T-l(T)+1)$ ,  $p_2 \alpha_2(T-l(T)+1)$ ,  $p_1 \alpha_1(T-l(T)+2)$ ,  $p_2 \alpha_2(T-l(T)+2)$ ,  $p_1 \alpha_1(T-l(T)+3)$ ,  $p_2 \alpha_2(T-l(T)+3)$ ,  $\dots$ ,  $p_1 \alpha_1(T)$ ,  $p_2 \alpha_2(T)$  (As we have mentioned before, the order of the Whittle indices alternates between the two classes because  $l(T)+1 \leq l_{\max}+1$ ). Moreover, the states of the users' proportion  $(1-p_1)\alpha_1(t)$  and  $(1-p_2)\alpha_2(t)$ ; which are scheduled but they don't transit to the state 1 with respect to their classes; will be increased by one. Leveraging the above results, we provide the decreasing order of all users' proportions according to the Whittle index value depending on two cases of  $\beta(t)$ . If  $\beta(T) = 0$ , then the smallest state's value among the users' proportions  $(1-p_1)\alpha_1(t)$  and  $(1-p_2)\alpha_2(t)$  at time  $T+1$  is  $l(T)+1$ . Hence, their Whittle index values will be higher than  $w_2(l(T)+1)$ , and consequently, they will be higher than those of users' proportion of  $\gamma(T) p_2 \alpha_2(T-l(T))$  at state  $l(T)+1$  in class 2.

If  $\beta(T) \neq 1$ , the smallest state value among the users' proportions  $(1-p_1)\alpha_1(t)$  and  $(1-p_2)\alpha_2(t)$  at time  $T+1$  is respectively  $l(T)+1$  and  $l(T)+2$ . Then, their Whittle index values will be higher than  $w_1(l(T)+1)$  ( $w_1(l(T)+1) < w_2(l(T)+2)$ ) as the alternation condition is satisfied from 1 until  $l_{\max}+1$ ). Consequently, their Whittle index values will be higher than the Whittle index of users' proportion  $\beta(T) p_1 \alpha_1(T-l(T))$  at state  $l(T)+1$  in class 1.

Thus, the decreasing order of all users' proportions according to the Whittle index value whatever the value of  $\beta(T)$  at  $T+1$  is:  $(1-p_1)\alpha_1(t)$ ,  $(1-p_2)\alpha_2(t)$ ,  $\beta(T) p_1 \alpha_1(T-l(T))$ ,  $\gamma(T) p_2 \alpha_2(T-l(T))$ ,  $p_1 \alpha_1(T-l(T)+1)$ ,  $p_2 \alpha_2(T-l(T)+1)$ ,  $p_1 \alpha_1(T-l(T)+2)$ ,  $p_2 \alpha_2(T-l(T)+2)$ ,  $p_1 \alpha_1(T-l(T)+3)$ ,  $p_2 \alpha_2(T-l(T)+3)$ ,  $\dots$ ,  $p_1 \alpha_1(T)$ ,  $p_2 \alpha_2(T)$ . As we have that  $(1-p_1)\alpha_1(t) + (1-p_2)\alpha_2(t) \leq \alpha$ , then surely the thresholds at time  $T+1$  in class 1 and in class 2 are less than the state of the users' proportion  $\beta(T) p_1 \alpha_1(T-l(T))$  and  $\gamma(T) p_2 \alpha_2(T-l(T))$  respectively. Therefore, there exists  $l_1(T+1)$ ,  $l_2(T+1)$ ,  $\beta(T+1)$  and  $\gamma(T+1)$  such that  $0 < \beta(T+1) \leq 1$  and

$\gamma(T + 1) = 1$ , or  $\beta(T + 1) = 1$  and  $0 < \gamma(T + 1) \leq 1$ :

$$\begin{aligned} & \sum_{j=1}^{l_1(T+1)-1} p_1 \alpha_1(T + 1 - j) + \sum_{j=1}^{l_2(T+1)-1} p_2 \alpha_2(T + 1 - j) \\ & + \beta(T + 1) p_1 \alpha_1(T + 1 - l_1(T + 1)) + \gamma(T) p_2 \alpha_2(T + 1 - l_2(T + 1)) = 1 - \alpha \end{aligned} \quad (\text{C.52})$$

Now we prove by contradiction that  $\max(l_1(T + 1), l_2(T + 1)) \leq T_{max}$ .

We prove first that  $l_2(T + 1)$  is greater than  $l_1(T + 1)$ .

As  $w_2(l_1(T + 1)) < w_1(l_1(T + 1))$ , that means according to lemma C.6,  $l_2(T + 1)$  is greater than  $l_1(T + 1)$  ( $w_2(l_1(T + 1)) < w_2(l_2(T + 1))$ ).

Reasoning by contradiction, if  $l_2(T + 1) > T_{max}$ , then we distinguish between two cases:

– First case: If  $\beta(T + 1) = 1$ :

we have that  $w_1(T_{max}) < w_2(l_2(T + 1))$  ( $w_1(T_{max}) < w_2(T_{max} + 1)$ ) as the alternation condition is satisfied in  $[1, T_{max} + 1]$ , i.e., according to lemma C.6, we have that  $l_{max} \leq l_1(T + 1)$ . Hence, according to lemmas 5.2 and 5.3, we have that:

$$\begin{aligned} & \sum_{j=1}^{l_1(T+1)-1} p_1 \alpha_1(T + 1 - j) + \sum_{j=1}^{l_2(T+1)-1} p_2 \alpha_2(T + 1 - j) \\ & + \beta(T + 1) p_1 \alpha_1(T + 1 - l_1(T + 1)) + \gamma(T + 1) p_2 \alpha_2(T + 1 - l_2(T + 1)) \\ & = 1 - \alpha \geq \sum_{j=1}^{T_{max}} p_1 \alpha_1(T + 1 - j) + \sum_{j=1}^{T_{max}} p_2 \alpha_2(T + 1 - j) > T_{max} p_2 \alpha \geq 1 - \alpha \end{aligned} \quad (\text{C.53})$$

Therefore we end up with:

$$1 - \alpha > 1 - \alpha \quad (\text{C.54})$$

Hence, the assumption that  $l_2(T + 1) > T_{max}$  leads us to an illogical statement. Consequently, the hypothesis of  $l_2(T + 1) > l_{max}$  is not valid for the first case.

– Second case: If  $\beta(T + 1) < 1$ :

Then we have that  $\gamma(T + 1) = 1$ . Therefore, all users at state  $l_2(T + 1)$  in class 2 are in the proportion  $1 - \alpha$  with the smallest Whittle index values. However, there are users in state  $l_1(T + 1)$  in class 1 of the  $\alpha$  proportion with the highest Whittle index values. In other words,  $w_1(l_1(T + 1)) \geq w_2(l_2(T + 1)) > w_1(T_{max})$ . This means that  $l_1(T + 1) > T_{max}$ . Therefore, according to lemmas 5.2 and

5.3:

$$\begin{aligned}
 & \sum_{j=1}^{l_1(T+1)-1} p_1 \alpha_1(T+1-j) + \sum_{j=1}^{l_2(T+1)-1} p_2 \alpha_2(T+1-j) \\
 & \quad + \beta(T+1)p_1 \alpha_1(T+1-l_1(T+1)) + \gamma(T+1)p_2 \alpha_2(T+1-l_2(T+1)) \\
 & = 1 - \alpha \geq \sum_{j=1}^{T_{max}} p_1 \alpha_1(T+1-j) + \sum_{j=1}^{T_{max}} p_2 \alpha_2(T+1-j) > T_{max} p_2 \alpha \geq 1 - \alpha
 \end{aligned} \tag{C.55}$$

Hence,

$$1 - \alpha > 1 - \alpha \tag{C.56}$$

Therefore, the hypothesis of  $l_2(T+1) > T_{max}$  is not valid for the second case.

Consequently, we have that  $l_2(T+1) \leq T_{max}$ , i.e.  $\max(l_1(T+1), l_2(T+1)) \leq T_{max}$ . Then, according to Proposition 5.10, there exists  $l(T+1)$ , and  $\gamma(T+1)$  and  $\beta(T+1)$  such that:

$$\begin{aligned}
 & \sum_{j=1}^{l(T+1)-1} p_1 \alpha_1(T+1-j) + \sum_{j=1}^{l(T+1)-1} p_2 \alpha_2(T+1-j) \\
 & \quad + \beta(T+1)p_1 \alpha_1(T+1-l(T+1)) + \gamma(T+1)p_2 \alpha_2(T+1-l(T+1)) = 1 - \alpha
 \end{aligned} \tag{C.57}$$

where  $\beta(T+1) = 0$  and  $0 < \gamma(T+1) \leq 1$ , or  $0 < \beta(T+1) \leq 1$  and  $\gamma(T+1) = 1$ .

To conclude, we have proved by induction, that for all  $T \geq T_0$ , there exists  $l(T)$ ,  $\beta(T)$  and  $\gamma(T)$ , such that:

$$\sum_{j=1}^{l(T)-1} p_1 \alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2 \alpha_2(T-j) + \beta(T)p_1 \alpha_1(T-l(T)) + \gamma(T)p_2 \alpha_2(T-l(T)) = 1 - \alpha \tag{C.58}$$

where  $\beta(T) = 0$  and  $0 < \gamma(T) \leq 1$ , or  $0 < \beta(T) \leq 1$  and  $\gamma(T) = 1$ , which concludes our proof.

## C.13 Proof of Proposition 5.12

We proceed by the same method used to prove the Proposition 5.11.

We consider at time  $T$ :

$$\sum_{j=1}^{l(T)-1} p_1 \alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2 \alpha_2(T-j) + \beta(T)p_1 \alpha_1(T-l(T)) + \gamma(T)p_2 \alpha_2(T-l(T)) = 1 - \alpha, \tag{C.59}$$

where  $\beta(T) = 0$  and  $0 \leq \gamma(T) < 1$ , or  $0 \leq \beta(T) < 1$  and  $\gamma(T) = 1$ . Among the users' proportion scheduled  $\alpha$ , exactly  $p_1\alpha_1(T)$  and  $p_2\alpha_2(T)$  will go to state one for each classes, and  $(1 - p_1)\alpha_1(T)$  and  $(1 - p_2)\alpha_2(T)$  will go to the next state.

For the other users for which the action taken is passive, their states will be increased by one, then the decreasing order according to the Whittle index value at the next time slot is  $\beta(T)p_1\alpha_1(T - l(T)), \gamma(T)p_2\alpha_2(T - l(T)), p_1\alpha_1(T - l(T) + 1), p_2\alpha_2(T - l(T) + 1), p_1\alpha_1(T - l(T) + 2), p_2\alpha_2(T - l(T) + 2) \cdots p_1\alpha_1(T), p_2\alpha_2(T)$  (As we said before that the order based on the value of the Whittle indices, alternate between the two classes from state 1 to  $l(T) \leq l_{\max} + 1$ ). Moreover, the users' proportion scheduled  $(1 - p_1)\alpha_1(T)$  and  $(1 - p_2)\alpha_2(T)$  will be at states that have Whittle index values higher than those of  $\beta(T)p_1\alpha_1(T - l(T))$  and  $\gamma(T)p_2\alpha_2(T - l(T))$  (as we have explained in the proof of Proposition 5.11).

Hence, the global decreasing order according to the Whittle index value is  $(1 - p_1)\alpha_1(T), (1 - p_1)\alpha_2(T), \beta(T)p_1\alpha_1(T - l(T)), \gamma(T)p_2\alpha_2(T - l(T)), p_1\alpha_1(T - l(T) + 1), p_2\alpha_2(T - l(T) + 1), p_1\alpha_1(T - l(T) + 2), p_2\alpha_2(T - l(T) + 2) \cdots p_1\alpha_1(T), p_2\alpha_2(T)$ .

Providing that  $(1 - p_1)\alpha_1(t) + (1 - p_2)\alpha_2(t) \leq \alpha$ , then at time  $T + 1$ :

$$\sum_{j=1}^{l(T)-1} p_1\alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2\alpha_2(T-j) + \beta(T)p_1\alpha_1(T-l(T)) + \gamma(T)p_2\alpha_2(T-l(T)) + p_1\alpha_1(T) + p_2\alpha_2(T) \geq 1 - \alpha \quad (\text{C.60})$$

Then, there exists  $\beta = 0$  and  $0 < \gamma \leq 1$ , or  $0 < \beta \leq 1$  and  $\gamma = 1$ , and sub-set  $\{\alpha_1(T), \alpha_2(T), \alpha_1(T - 1), \alpha_2(T - 1) \cdots \alpha_1(T - m), \alpha_2(T - m)\} \subset \{\alpha_1(T - l(T)), \alpha_2(T - l(T)), \alpha_1(T - l(T) + 1), \alpha_2(T - l(T) + 1), \alpha_1(T - l(T) + 2), \alpha_2(T - l(T) + 2) \cdots \alpha_1(T), \alpha_2(T)\}$ , such that:

$$\sum_{j=1}^{(m+1)-1} p_1\alpha_1(T+1-j) + \sum_{j=1}^{(m+1)-1} p_2\alpha_2(T+1-j) + \beta p_1\alpha_1(T+1-(m+1)) + \gamma p_2\alpha_2(T+1-(m+1)) = 1 - \alpha \quad (\text{C.61})$$

Indeed,  $m + 1$  is effectively  $l(T + 1)$ ,  $\beta = \beta(T + 1)$ ,  $\gamma = \gamma(T + 1)$ , and the elements of the set  $\{\alpha_1(T), \alpha_1(T - 1), \cdots \alpha_1(T - m)\} \cup \{\alpha_1(T + 1)\}$  and the set  $\{\alpha_2(T), \alpha_2(T - 1), \cdots, \alpha_2(T - m)\} \cup \{\alpha_2(T + 1)\}$  are exactly the elements of the vectors  $A_1(T + 1)$  and  $A_2(T + 1)$  respectively. Given that  $\{\alpha_1(T), \alpha_1(T - 1), \cdots \alpha_1(T - m)\}$  and  $\{\alpha_2(T), \alpha_2(T - 1), \cdots, \alpha_2(T - m)\}$  are included in the set of elements of the vector  $A_1(T)$  and  $A_2(T)$  respectively, then for  $k = 1, 2$ , all the elements of the vector  $A_k(T + 1)$  except  $\alpha_k(T + 1)$  belong to the elements of vector  $A_k(T)$ .

## C.14 Proof of Proposition 5.13

According to Proposition 5.11, the elements of the vectors  $A_1(T)$  and  $A_2(T)$  satisfy:

$$\sum_{j=1}^{l(T)-1} p_1\alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2\alpha_2(T-j) + \beta(T)p_1\alpha_1(T-l(T)) + \gamma(T)p_2\alpha_2(T-l(T)) = 1 - \alpha, \quad (\text{C.62})$$

where  $0 < \beta(T) \leq 1$  and  $\gamma(T) = 1$ , or  $\beta(T) = 0$  and  $0 < \gamma(T) \leq 1$ . We distinguish between two cases depending on the values of  $\beta$  and  $\gamma$  (we drop the index  $T$  on  $\beta(T)$  and

$\gamma(T)$  to ease the notation):

- First case:  $0 < \beta \leq 1$ , and  $\gamma = 1$ :

Hence:

$$\sum_{j=1}^{l(T)-1} p_1 \alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2 \alpha_2(T-j) + \beta(T) p_1 \alpha_1(T-l(T)) + p_2 \alpha_2(T-l(T)) = 1 - \alpha \quad (\text{C.63})$$

Our aim is to derive the expression of  $\alpha_k(T+1)$  for class 1 and class 2. Among the users' proportion scheduled  $\alpha$ , exactly  $p_1 \alpha_1(T)$  and  $p_2 \alpha_2(T)$  will go to state one for each class, and the rest will go to the next state. Hence:

$$\alpha_1(T+1) = (1 - p_1) \alpha_1(T) + B_1(T) \quad (\text{C.64})$$

$$\alpha_2(T+1) = (1 - p_2) \alpha_2(T) + B_2(T) \quad (\text{C.65})$$

such that  $B_1(T) + B_2(T) = p_1 \alpha_1(T) + p_2 \alpha_2(T)$ .

At time  $T+1$ , the decreasing order according to the Whittle index value is  $(1 - p_1) \alpha_1(T)$ ,  $(1 - p_2) \alpha_2(T)$ ,  $\beta p_1 \alpha_1(T - l(T))$ ,  $p_2 \alpha_2(T - l(T))$ ,  $p_1 \alpha_1(T - l(T) + 1)$ ,  $p_2 \alpha_2(T - l(T) + 1)$ ,  $p_1 \alpha_1(T - l(T) + 2)$ ,  $p_2 \alpha_2(T - l(T) + 2)$ ,  $\dots$ ,  $p_1 \alpha_1(T)$ ,  $p_2 \alpha_2(T)$ .

In order to get  $B_1(T)$  and  $B_2(T)$ , we sum the users' proportions at different states starting from the users' proportion  $\beta p_1 \alpha_1(T - l(T))$  following the decreasing order of the Whittle index until we get the sum that equals to  $p_1 \alpha_1(T) + p_2 \alpha_2(T)$ . We distinguish between six sub-cases and for each sub-case, we prove that  $\alpha_k(T+1)$  is surely between two elements of the vector  $A_k(T)$ . In fact, if we prove it just for one class, the result will be true for the other one, since  $\alpha_1(T) + \alpha_2(T) = \alpha$  for all  $T$ . In the following, we derive the expression of  $\alpha_k(T+1)$  for  $k = 1, 2$ , in function of the elements of the vector  $A_1(T)$  and  $A_2(T)$  and we show that  $\alpha_1(T+1)$  is surely between two elements of the vector  $A_1(T)$ .

1) If  $p_1 \alpha_1(T) + p_2 \alpha_2(T) \leq p_1 \beta \alpha_1(T - l(T))$ :

In this case  $p_1 \alpha_1(T) + p_2 \alpha_2(T)$  is less than  $p_1 \beta \alpha_1(T - l(T))$ . Therefore, we will take a proportion of users from  $p_1 \beta \alpha_1(T - l(T))$  that equals to  $p_1 \alpha_1(T) + p_2 \alpha_2(T)$  denoted by  $C$ . This users' proportion exactly equals to  $B_1(T) + B_2(T)$  that we add to  $(1 - p_1) \alpha_1(T)$  and  $(1 - p_2) \alpha_2(T)$ . Thus,  $B_1(T) + B_2(T) = C$ . However, since all the users of the proportion  $C$  belong to  $p_1 \beta \alpha_1(T - l(T))$ , then  $C$  contains only the users of the class 1. Consequently,  $B_1(T) = C$  and  $B_2(T) = 0$ . Hence:

$$\alpha_2(T+1) = (1 - p_2) \alpha_2(T) \quad (\text{C.66})$$

As  $\alpha_1(T+1) + \alpha_2(T+1) = \alpha$ , then:

$$\alpha_1(T+1) = \alpha - \alpha_2(T+1) \quad (\text{C.67})$$

Now we find the upper bound of  $\alpha_2(T) - \alpha_2(T+1)$ :

$$\alpha_2(T) - \alpha_2(T+1) = p_2 \alpha(T) \quad (\text{C.68})$$

$$\leq \beta \alpha_1(T - l(T)) p_1 - \alpha_1(T) p_1 \quad (\text{C.69})$$

$$\leq p_1 (\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.70})$$

$$= p_1 (\alpha_2(T) - \alpha_2(T - l(T))) \quad (\text{C.71})$$

The first inequality comes from the fact that  $p_1\alpha_1(T) + p_2\alpha_2(T) \leq p_1\beta\alpha_1(T - l(T))$  and the second one comes from the fact that  $\beta \leq 1$ .

Given that  $\alpha_2(i) - \alpha_2(j) = \alpha_1(j) - \alpha_1(i)$  for all integers  $i$  and  $j$ , thus:

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.72})$$

Moreover, we have that  $\alpha_1(T + 1) - \alpha_1(T) \geq 0$  because  $\alpha_2(T + 1) - \alpha_2(T) \leq 0$ . Therefore,  $\alpha_1(T) \leq \alpha_1(T + 1)$ . On the other hands, as  $p_1(\alpha_1(T - l(T)) - \alpha_1(T)) \geq \alpha_1(T + 1) - \alpha_1(T) \geq 0$  then  $\alpha_1(T - l(T)) - \alpha_1(T) \geq \alpha_1(T + 1) - \alpha_1(T)$ . This means that  $\alpha_1(T - l(T)) \geq \alpha_1(T + 1)$ . Consequently, we end up with:

$$\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T)) \quad (\text{C.73})$$

2) If  $\beta\alpha_1(T - l(T))p_1 \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \beta\alpha_1(T - l(T))p_1 + \alpha_2(T - l(T))p_2$ :  
Hence:

$$\alpha_1(T + 1) = (1 - p_1)\alpha_1(T) + \beta p_1\alpha_1(T - l(T)) \quad (\text{C.74})$$

$$\alpha_2(T + 1) = \alpha - \alpha_1(T + 1) \quad (\text{C.75})$$

Then:

$$\alpha_1(T + 1) - \alpha_1(T) = \beta p_1\alpha_1(T - l(T)) - p_1\alpha_1(T) \quad (\text{C.76})$$

$$\leq p_1(\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.77})$$

On the other hand, we have according to the right inequality of sub-case's assumption:

$$\alpha_1(T + 1) - \alpha_1(T) = \beta p_1\alpha_1(T - l(T)) - p_1\alpha_1(T) \quad (\text{C.78})$$

$$\geq p_2\alpha_2(T) - p_2\alpha_2(T - l(T)) \quad (\text{C.79})$$

$$= p_2(\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.80})$$

Hence :

$$p_2(\alpha_1(T - l(T)) - \alpha_1(T)) \leq \alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.81})$$

Knowing that  $p_2 < p_1$ , the later inequalities imply that  $\alpha_1(T - l(T)) - \alpha_1(T) \geq 0$ . As a result we have that:

$$\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T)) \quad (\text{C.82})$$

And

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.83})$$

3) If  $\beta\alpha_1(T - l(T))p_1 + \alpha_2(T - l(T))p_2 \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \beta\alpha_1(T - l(T))p_1 + \alpha_2(T - l(T))p_2 + p_1\alpha_1(T - l(T) + 1)$ :

Hence:

$$\alpha_2(T + 1) = (1 - p_2)\alpha_2(T) + p_2\alpha_2(T - l(T)) \quad (\text{C.84})$$

$$\alpha_1(T + 1) = \alpha - \alpha_2(T + 1) \quad (\text{C.85})$$

Therefore:

$$\alpha_2(T+1) - \alpha_2(T) = p_2(\alpha_2(T-l(T)) - \alpha_2(T)) \quad (\text{C.86})$$

And:

$$\alpha_1(T) - \alpha_1(T+1) = p_2(\alpha_1(T) - \alpha_1(T-l(T))) \quad (\text{C.87})$$

This means that if  $\alpha_1(T) \leq \alpha_1(T+1)$ :

$$\alpha_1(T) \leq \alpha_1(T+1) \leq \alpha_1(T-l(T)) \quad (\text{C.88})$$

And

$$\alpha_1(T+1) - \alpha_1(T) \leq p_1(\alpha_1(T-l(T)) - \alpha_1(T)) \quad (\text{C.89})$$

If  $\alpha_1(T+1) \leq \alpha_1(T)$ :

$$\alpha_1(T-l(T)) \leq \alpha_1(T+1) \leq \alpha_1(T) \quad (\text{C.90})$$

And

$$\alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T))) \quad (\text{C.91})$$

4) If  $\beta\alpha_1(T-l(T))p_1 + \alpha_2(T-l(T))p_2 + p_1\alpha_1(T-l(T)+1) \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \beta\alpha_1(T-l(T))p_1 + \alpha_2(T-l(T))p_2 + p_1\alpha_1(T-l(T)+1) + p_2\alpha_2(T-l(T)+1)$  :

Hence:

$$\alpha_1(T+1) = (1-p_1)\alpha_1(T) + p_1\beta\alpha_1(T-l(T)) + p_1\alpha_1(T-l(T)+1) \quad (\text{C.92})$$

$$\alpha_2(T+1) = \alpha - \alpha_1(T+1) \quad (\text{C.93})$$

Therefore:

$$\alpha_1(T+1) - \alpha_1(T) = -p_1\alpha_1(T) + p_1\beta\alpha_1(T-l(T)) + p_1\alpha_1(T-l(T)+1) \quad (\text{C.94})$$

$$\quad (\text{C.95})$$

According to the left inequality of the assumption of this case, we have that:

$$\alpha_1(T+1) - \alpha_1(T) \leq p_2\alpha_2(T) - p_2\alpha_2(T-l(T)) \quad (\text{C.96})$$

$$= p_2(\alpha_1(T-l(T)) - \alpha_1(T)) \quad (\text{C.97})$$

On the other hand, we have that:

$$\alpha_1(T+1) - \alpha_1(T) = -p_1\alpha_1(T) + p_1\beta\alpha_1(T-l(T)) + p_1\alpha_1(T-l(T)+1) \quad (\text{C.98})$$

$$\geq p_1(\alpha_1(T-l(T)+1) + \alpha_1(T)) \quad (\text{C.99})$$

$$\quad (\text{C.100})$$

Hence:

$$p_1(\alpha_1(T-l(T)+1) - \alpha_1(T)) \leq \alpha_1(T+1) - \alpha_1(T) \leq p_2(\alpha_1(T-l(T)) - \alpha_1(T)) \quad (\text{C.101})$$

Thus:

If  $\alpha_1(T) \leq \alpha_1(T+1)$ :

$$\alpha_1(T) \leq \alpha_1(T+1) \leq \alpha_1(T-l(T)) \quad (\text{C.102})$$

And

$$\alpha_1(T+1) - \alpha_1(T) \leq p_2(\alpha_1(T-l(T)) - \alpha_1(T)) \quad (\text{C.103})$$

$$\leq p_1(\alpha_1(T-l(T)) - \alpha_1(T)) \quad (\text{C.104})$$

If  $\alpha_1(T+1) \leq \alpha_1(T)$ :

$$\alpha_1(T-l(T)+1) \leq \alpha_1(T+1) \leq \alpha_1(T) \quad (\text{C.105})$$

And

$$\alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \quad (\text{C.106})$$

5) If there exists  $m \geq 1$  such that:

$$\begin{aligned} & \beta\alpha_1(T-l(T))p_1 + \alpha_2(T-l(T))p_2 + p_1\alpha_1(T-l(T)+1) + \cdots + p_1\alpha_1(T-l(T)+m) + \\ & p_2\alpha_2(T-l(T)+m) \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \beta\alpha_1(T-l(T))p_1 + \alpha_2(T-l(T))p_2 + \\ & p_1\alpha_1(T-l(T)+1) + \cdots + p_1\alpha_1(T-l(T)+m) + p_2\alpha_2(T-l(T)+m) + p_1\alpha_1(T-l(T)+m+1) : \end{aligned}$$

This means that:

$$\alpha_2(T+1) = (1-p_2)\alpha_2(T) + p_2\alpha_2(T-l(T)) + \cdots + p_2\alpha_2(T-l(T)+m) \quad (\text{C.107})$$

$$\alpha_1(T+1) = \alpha_1(T) - \alpha_2(T+1) \quad (\text{C.108})$$

We have that:

$$\begin{aligned} \alpha_2(T+1) - \alpha_2(T) &= -p_2\alpha_2(T) + p_2\alpha_2(T-l(T)) + p_2\alpha_1(T-l(T)+1) \\ &+ \cdots + p_2\alpha_2(T-l(T)+m) \end{aligned} \quad (\text{C.109})$$

$$\geq p_2(\alpha_2(T-l(T)+1) - \alpha_2(T)) \quad (\text{C.110})$$

On the other hand:

$$\begin{aligned} \alpha_2(T+1) - \alpha_2(T) &= -p_2\alpha_2(T) + p_2\alpha_2(T-l(T)) + p_2\alpha_2(T-l(T)+1) \\ &+ \cdots + p_2\alpha_2(T-l(T)+m) \end{aligned} \quad (\text{C.111})$$

$$\leq p_1\alpha_1(T) - \beta p_1\alpha_1(T-l(T)) - \sum_{i=1}^m p_1\alpha_1(T-l(T)+i) \quad (\text{C.112})$$

$$\leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \quad (\text{C.113})$$

$$= p_1(\alpha_2(T-l(T)+1) - \alpha_2(T)) \quad (\text{C.114})$$

Thus:

$$p_2(\alpha_1(T) - \alpha_1(T-l(T)+1)) \leq \alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \quad (\text{C.115})$$

Therefore:

$$\alpha_1(T-l(T)+1) \leq \alpha_1(T+1) \leq \alpha_1(T) \quad (\text{C.116})$$

And:

$$\alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \quad (\text{C.117})$$

6) If there exists  $m \geq 1$  such that:

$$\begin{aligned} & \beta\alpha_1(T-l(T))p_1 + \alpha_2(T-l(T))p_2 + p_1\alpha_1(T-l(T)+1) + p_2\alpha_2(T-l(T)+1) + \cdots + \\ & p_1\alpha_1(T-l(T)+m) + p_2\alpha_2(T-l(T)+m) + p_1\alpha_1(T-l(T)+m+1) \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \\ & \beta\alpha_1(T-l(T))p_1 + \alpha_2(T-l(T))p_2 + p_1\alpha_1(T-l(T)+1) + \cdots + p_1\alpha_1(T-l(T)+m) + \\ & p_2\alpha_2(T-l(T)+m) + p_1\alpha_1(T-l(T)+m+1) + p_2\alpha_2(T-l(T)+m+1) : \end{aligned}$$

Hence:

$$\alpha_1(T+1) = (1-p_1)\alpha_1(T) + p_1\beta\alpha_1(T-l(T)) + \cdots + p_1\alpha_1(T-l(T)+m+1) \quad (\text{C.118})$$

$$\alpha_2(T+1) = \alpha - \alpha_1(T+1) \quad (\text{C.119})$$

We have that:

$$\begin{aligned} \alpha_1(T+1) - \alpha_1(T) &= -p_1\alpha_1(T) + p_1\beta\alpha_1(T-l(T)) + p_1\alpha_1(T-l(T)+1) \\ &+ \cdots + p_1\alpha_1(T-l(T)+m+1) \end{aligned} \quad (\text{C.120})$$

$$\geq p_1(\alpha_1(T-l(T)+1) - \alpha_1(T)) \quad (\text{C.121})$$

On the other hand:

$$\begin{aligned} \alpha_1(T+1) - \alpha_1(T) &= -p_1\alpha_1(T) + p_1\beta\alpha_1(T-l(T)) + p_1\alpha_1(T-l(T)+1) \\ &+ \cdots + p_1\alpha_1(T-l(T)+m+1) \end{aligned} \quad (\text{C.122})$$

$$\leq p_2\alpha_2(T) - \sum_{i=0}^m p_2\alpha_2(T-l(T)+i) \quad (\text{C.123})$$

$$\leq p_2(\alpha_2(T) - \alpha_2(T-l(T)+1)) \quad (\text{C.124})$$

$$= p_2(\alpha_1(T-l(T)+1) - \alpha_1(T)) \quad (\text{C.125})$$

Thus:

$$p_1(\alpha_1(T-l(T)+1) - \alpha_1(T)) \leq \alpha_1(T+1) - \alpha_1(T) \leq p_2(\alpha_1(T-l(T)+1) - \alpha_1(T)) \quad (\text{C.126})$$

Therefore:

$$\alpha_1(T-l(T)+1) \leq \alpha_1(T+1) \leq \alpha_1(T) \quad (\text{C.127})$$

And:

$$\alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \quad (\text{C.128})$$

- Second case:  $\beta = 0$  and  $0 < \gamma \leq 1$ :

Hence, we have that:

$$\sum_{j=1}^{l(T)-1} p_1\alpha_1(T-j) + \sum_{j=1}^{l(T)-1} p_2\alpha_2(T-j) + \gamma p_2\alpha_2(T-l(T)) = 1 - \alpha \quad (\text{C.129})$$

Then, at time  $T+1$ , the decreasing order according to the Whittle index value is  $(1-p_1)\alpha_1(T)$ ,  $(1-p_2)\alpha_2(T)$ ,  $\gamma p_2\alpha_2(T-l(T))$ ,  $p_1\alpha_1(T-l(T)+1)$ ,  $p_2\alpha_2(T-l(T)+1)$ ,  $p_1\alpha_1(T-l(T)+2)$ ,  $p_2\alpha_2(T-l(T)+2)$ ,  $\cdots$ ,  $p_1\alpha_1(T)$ ,  $p_2\alpha_2(T)$ . In order to obtain  $B_1(T)$  and  $B_2(T)$ , we sum the users' proportions at different states starting from the users' proportion  $\gamma p_2\alpha_2(T-l(T))$  following the decreasing order of the Whittle

index until we get the sum that equals to  $p_1\alpha_1(T) + p_2\alpha_2(T)$ . For this case, we distinguish between five sub-cases, and for each sub-case, we prove that  $\alpha_1(T + 1)$  is surely between two elements of the vector  $A_1(T)$ .

1) If  $p_1\alpha_1(T) + p_2\alpha_2(T) \leq \gamma\alpha_2(T - l(T))p_2$ :

Hence:

$$\alpha_1(T + 1) = (1 - p_1)\alpha_1(T) \quad (\text{C.130})$$

$$\alpha_2(T + 1) = \alpha - \alpha_1(T + 1) \quad (\text{C.131})$$

We have that:

$$\alpha_1(T) - \alpha_1(T + 1) = p_1\alpha_1(T) \quad (\text{C.132})$$

$$\leq \gamma\alpha_2(T - l(T))p_2 - \alpha_2(T)p_2 \quad (\text{C.133})$$

$$\leq p_2(\alpha_2(T - l(T)) - \alpha_2(T)) \quad (\text{C.134})$$

$$= p_2(\alpha_1(T) - \alpha_1(T - l(T))) \quad (\text{C.135})$$

Thus:

$$\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T))) \quad (\text{C.136})$$

And:

$$\alpha_1(T - l(T)) \leq \alpha_1(T + 1) \leq \alpha_1(T) \quad (\text{C.137})$$

2) If  $\gamma\alpha_2(T - l(T))p_2 \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \gamma\alpha_2(T - l(T))p_2 + \alpha_1(T - l(T) + 1)p_1$

Consequently:

$$\alpha_2(T + 1) = (1 - p_2)\alpha_2(T) + \gamma p_2 \alpha_2(T - l(T)) \quad (\text{C.138})$$

$$\alpha_1(T + 1) = \alpha - \alpha_2(T + 1) \quad (\text{C.139})$$

Hence:

$$\alpha_2(T + 1) - \alpha_2(T) = -p_2\alpha_2(T) + \gamma p_2 \alpha_2(T - l(T)) \quad (\text{C.140})$$

$$\leq p_2(\alpha_2(T - l(T)) - \alpha_2(T)) \quad (\text{C.141})$$

$$(\text{C.142})$$

On the other hand, according to the right inequality of the assumption of this case, we have that:

$$\alpha_2(T + 1) - \alpha_2(T) = -p_2\alpha_2(T) + \gamma p_2 \alpha_2(T - l(T)) \quad (\text{C.143})$$

$$\geq p_1(\alpha_1(T) - \alpha_1(T - l(T) + 1)) \quad (\text{C.144})$$

$$= p_1(\alpha_2(T - l(T) + 1) - \alpha_2(T)) \quad (\text{C.145})$$

That means:

$$p_1(\alpha_2(T - l(T) + 1) - \alpha_2(T)) \leq \alpha_2(T + 1) - \alpha_2(T) \leq p_2(\alpha_2(T - l(T)) - \alpha_2(T)) \quad (\text{C.146})$$

i.e.

$$p_1(\alpha_1(T) - \alpha_1(T - l(T) + 1)) \leq \alpha_1(T) - \alpha_1(T + 1) \leq p_2(\alpha_1(T) - \alpha_1(T - l(T))) \quad (\text{C.147})$$

Therefore:

If  $\alpha_1(T) \leq \alpha_1(T + 1)$ :

$$\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1) \quad (\text{C.148})$$

And:

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T)) \quad (\text{C.149})$$

If  $\alpha_1(T + 1) \leq \alpha_1(T)$ :

$$\alpha_1(T - l(T)) \leq \alpha_1(T + 1) \leq \alpha_1(T) \quad (\text{C.150})$$

And:

$$\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T))) \quad (\text{C.151})$$

3) If  $\gamma\alpha_2(T - l(T))p_2 + \alpha_1(T - l(T) + 1)p_1 \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \gamma\alpha_2(T - l(T))p_2 + \alpha_1(T - l(T) + 1)p_1 + p_2\alpha_2(T - l(T) + 1)$ .

Hence:

$$\alpha_1(T + 1) = (1 - p_1)\alpha_1(T) + p_1\alpha_1(T - l(T) + 1) \quad (\text{C.152})$$

$$\alpha_2(T + 1) = \alpha - \alpha_1(T + 1) \quad (\text{C.153})$$

We have that:

$$\alpha_1(T + 1) - \alpha_1(T) = p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T)) \quad (\text{C.154})$$

If  $\alpha_1(T) \leq \alpha_1(T + 1)$ :

$$\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1) \quad (\text{C.155})$$

And:

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T)) \quad (\text{C.156})$$

If  $\alpha_1(T + 1) \leq \alpha_1(T)$ :

$$\alpha_1(T - l(T) + 1) \leq \alpha_1(T + 1) \leq \alpha_1(T) \quad (\text{C.157})$$

And:

$$\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T) + 1)) \quad (\text{C.158})$$

4) If there exists  $m \geq 1$  such that:

$\gamma\alpha_2(T - l(T))p_2 + \dots + \alpha_1(T - l(T) + m)p_1 + p_2\alpha_2(T - l(T) + m) \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \gamma\alpha_2(T - l(T))p_2 + \dots + \alpha_1(T - l(T) + m)p_1 + p_2\alpha_2(T - l(T) + m) + p_1\alpha_1(T - l(T) + m + 1)$ :

Hence:

$$\alpha_2(T + 1) = (1 - p_2)\alpha_2(T) + p_2\gamma\alpha_2(T - l(T)) + \dots + p_2\alpha_2(T - l(T) + m) \quad (\text{C.159})$$

$$\alpha_1(T + 1) = \alpha - \alpha_2(T + 1) \quad (\text{C.160})$$

$$\begin{aligned} \alpha_2(T+1) - \alpha_2(T) &= -p_2\alpha_2(T) + p_2\gamma\alpha_2(T-l(T)) + p_2\alpha_2(T-l(T)+1) \\ &\quad + \cdots + p_2\alpha_2(T-l(T)+m) \end{aligned} \tag{C.161}$$

$$\geq p_2(\alpha_2(T-l(T)+1) - \alpha_2(T)) \tag{C.162}$$

On the other hand:

$$\begin{aligned} \alpha_2(T+1) - \alpha_2(T) &= -p_2\alpha_2(T) + p_2\gamma\alpha_2(T-l(T)) + p_2\alpha_2(T-l(T)+1) \\ &\quad + \cdots + p_2\alpha_2(T-l(T)+m) \end{aligned} \tag{C.163}$$

$$\leq p_1\alpha_1(T) - \sum_{i=1}^m p_1\alpha_1(T-l(T)+i) \tag{C.164}$$

$$\leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \tag{C.165}$$

$$= p_1(\alpha_2(T-l(T)+1) - \alpha_2(T)) \tag{C.166}$$

Thus:

$$p_2(\alpha_1(T) - \alpha_1(T-l(T)+1)) \leq \alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \tag{C.167}$$

Therefore:

$$\alpha_1(T-l(T)+1) \leq \alpha_1(T+1) \leq \alpha_1(T) \tag{C.168}$$

And:

$$\alpha_1(T) - \alpha_1(T+1) \leq p_1(\alpha_1(T) - \alpha_1(T-l(T)+1)) \tag{C.169}$$

5) If there exists  $m \geq 1$  such that:

$$\begin{aligned} \gamma\alpha_2(T-l(T))p_2 + \cdots + \alpha_1(T-l(T)+m)p_1 + p_2\alpha_2(T-l(T)+m) + p_1\alpha_1(T-l(T)+m+1) \\ \leq p_1\alpha_1(T) + p_2\alpha_2(T) \leq \gamma\alpha_2(T-l(T))p_2 + \cdots + \alpha_1(T-l(T)+m)p_1 + p_2\alpha_2(T-l(T)+m) \\ + p_1\alpha_1(T-l(T)+m+1) + p_2\alpha_2(T-l(T)+m+1): \end{aligned}$$

That implies that:

$$\alpha_1(T+1) = (1-p_1)\alpha_1(T) + \cdots + p_1\alpha_1(T-l(T)+m) + p_1\alpha_1(T-l(T)+m+1) \tag{C.170}$$

$$\alpha_2(T+1) = \alpha - \alpha_1(T+1) \tag{C.171}$$

$$\alpha_1(T+1) - \alpha_1(T) = -p_1\alpha_1(T) + p_1\alpha_1(T-l(T)+1) + \cdots + p_1\alpha_1(T-l(T)+m+1) \tag{C.172}$$

$$\geq p_1(\alpha_1(T-l(T)+1) - \alpha_1(T)) \tag{C.173}$$

On the other hand:

$$\begin{aligned} \alpha_1(T+1) - \alpha_1(T) &= -p_1\alpha_1(T) + p_1\alpha_1(T-l(T)+1) + \cdots + p_1\alpha_1(T-l(T)+m+1) \end{aligned} \tag{C.174}$$

$$\begin{aligned} &\leq p_2\alpha_2(T) - \gamma p_2\alpha_2(T-l(T)) - \sum_{i=1}^m p_2\alpha_2(T-l(T)+i) \end{aligned} \tag{C.175}$$

$$\leq p_2(\alpha_2(T) - \alpha_2(T-l(T)+1)) \tag{C.176}$$

$$= p_2(\alpha_1(T-l(T)+1) - \alpha_1(T)) \tag{C.177}$$

Thus:

$$p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T)) \leq \alpha_1(T + 1) - \alpha_1(T) \leq p_2(\alpha_1(T - l(T) + 1) - \alpha_1(T)) \quad (\text{C.178})$$

Therefore:

$$\alpha_1(T - l(T) + 1) \leq \alpha_1(T + 1) \leq \alpha_1(T) \quad (\text{C.179})$$

And:

$$\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T) + 1)) \quad (\text{C.180})$$

In conclusion, all these six sub-cases when  $\gamma = 1$  and  $0 < \beta \leq 1$ , plus the five sub-cases when  $\beta = 0$  and  $0 < \gamma \leq 1$ , can be summarized in four cases:

- 1)  $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T))$ , and  $\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T)) - \alpha_1(T))$ .
- 2)  $\alpha_1(T - l(T)) \leq \alpha_1(T + 1) \leq \alpha_1(T)$ , and  $\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T)))$ .
- 3)  $\alpha_1(T - l(T) + 1) \leq \alpha_1(T + 1) \leq \alpha_1(T)$ , and  $\alpha_1(T) - \alpha_1(T + 1) \leq p_1(\alpha_1(T) - \alpha_1(T - l(T) + 1))$ .
- 4)  $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1)$ , and  $\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T))$ .

Thus, the proof is concluded.

## C.15 Proof of Proposition 5.14

In all the proof, we consider that  $\epsilon \leq (l_2 - l_1) \frac{(1-p_1)^L}{1-(1-p_1)^L}$ .

Before tackling the proof, we give a brief insight about the procedure adopted to establish the desired result: We start by finding a given time denoted  $T_2 \geq T_\epsilon$  where  $\alpha_1(T_2)$  is less than  $l_1$ . Then, we show that  $\alpha_1(T_2), \dots, \alpha_1(T_2 + L)$  are strictly less than  $l_2$ . To that end, we start first by defining a relevant sequence  $u_n$  in function of  $\epsilon, l_1, l_2$  and  $p_1$  when  $n \in [0, L]$ . After that, we prove that  $u_n$  is increasing with  $n$  and strictly less than  $l_2$ . Next, we establish that  $u_n$  is an upper bound of  $\alpha_1(\cdot)$  in  $[T_2, T_2 + L]$ . More precisely, we show that  $\alpha_1(T_2 + n) \leq u_n$  for  $n \in [0, L]$ . For that purpose, we proceed with two following steps: The first one consists of deriving an inequality verified by two consecutive terms of the sequence  $\alpha_1(\cdot)$ , namely  $\alpha_1(T)$  and  $\alpha_1(T + 1)$  using the Proposition 5.13 given that  $T \geq T_\epsilon$ . As for the second step, we use essentially the aforementioned result to demonstrate by induction that  $u_n$  is indeed an upper bound of  $\alpha_1(T_2 + n)$ . Finally, based on these results, we show that there exists  $T_d$  such that  $\max A_1(T_d) < l_2$ .

To find a time  $T_2 \geq T_\epsilon$  such that  $\alpha_1(T_2)$  is less than  $l_1$ , we use the fact that  $\min A_1(t) \leq l_1$  for all  $t$ . At time  $T_\epsilon + L$ , we have the vector  $A_1(T_\epsilon + L) = (\alpha_1(T_\epsilon + L), \alpha_1(T_\epsilon + L - 1), \dots, \alpha_1(T_\epsilon + L - l(T_\epsilon + L)))$ . Providing that  $\min A_1(T_\epsilon + L) \leq l_1$ , then there exists an element from the vector  $A_1(T_\epsilon + L)$  less than  $l_1$  denoted by  $\alpha_1(T_2)$ . According to Proposition 5.11, we have for all  $T \geq T_0$ ,  $l(T) \leq l_{\max} = L$ , then  $l(T_\epsilon + L) \leq L$ . That is,  $T_2$  is greater than  $T_\epsilon$  since  $T_2 \geq T_\epsilon + L - l(T_\epsilon + L) \geq T_\epsilon$ . Therefore, we find an element of the

sequence  $\alpha_1(\cdot)$  at time  $T_2 \geq T_\epsilon$  such that  $\alpha_1(T_2) \leq l_1$ . To that extent, we are interested in proving that  $\alpha_1(T_2), \dots, \alpha_1(T_2 + L)$  are strictly less than  $l_2$ .

To do so, we define a sequence  $u_n$  which will constitute an upper bound of the function  $\alpha_1(T)$ .

**Definition C.1.** *We define a sequence  $u_n$  by induction:*

$$\begin{cases} u_0 = l_1 & \text{if } n = 0 \\ u_{n+1} = p_1(l_2 + \epsilon) + (1 - p_1)u_n & \text{if } n > 0 \end{cases} \quad (\text{C.181})$$

Next, we prove that the  $L$  first terms of this sequence are strictly less than  $l_2$ . We detail this in the following.

**Lemma C.7.** *For  $n \in [0, L]$ ,  $u_n < l_2$*

*Proof.* In fact, the sequence  $u_n$  satisfies for all  $n$ :

$$u_n = \lambda(1 - p_1)^n + (l_2 + \epsilon) \quad (\text{C.182})$$

where  $\lambda = -(\epsilon + l_2 - l_1)$ .

$u_n$  is clearly increasing with  $n$ , then for all  $n \in [0, L]$ :

$$u_n \leq u_L = \lambda(1 - p_1)^L + (l_2 + \epsilon) = \epsilon(1 - (1 - p_1)^L) + l_2 - (l_2 - l_1)(1 - p_1)^L \quad (\text{C.183})$$

We have that:

$$\epsilon < (l_2 - l_1) \left( \frac{(1 - p_1)^L}{1 - (1 - p_1)^L} \right) \quad (\text{C.184})$$

Given that  $1 - (1 - p_1)^L \geq 0$ , then:

$$(1 - (1 - p_1)^L)\epsilon < (l_2 - l_1)(1 - p_1)^L \quad (\text{C.185})$$

$$(1 - (1 - p_1)^L)\epsilon + l_2 - (l_2 - l_1)(1 - p_1)^L < l_2 \quad (\text{C.186})$$

Therefore,  $u_L < l_2$ . □

Based on the lemma above, we prove that for any element of the set  $\{\alpha_1(T_2), \dots, \alpha_1(T_2 + L)\}$  must be less than  $u_L$ .

For that, we introduce a useful Lemma:

**Lemma C.8.** *If for  $T \in [T_2, T_2 + L - 1]$ , we have that:*

$$\alpha_1(T) \leq \alpha_1(T + 1) \quad (\text{C.187})$$

*Then, we have that:*

$$\alpha_1(T + 1) \leq p_1(l_2 + \epsilon) + (1 - p_1)\alpha_1(T) \quad (\text{C.188})$$

*Proof.* Before starting the proof, we recall that, according to the first result of Proposition 5.13, the four possible inequalities satisfied by  $\alpha_1(T)$ ,  $\alpha_1(T + 1)$ ,  $\alpha_1(T - l(T))$ ,  $\alpha_1(T - l(T) + 1)$  are:

$$\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T)) \quad (\text{C.189})$$

$$\alpha_1(T - l(T)) \leq \alpha_1(T + 1) \leq \alpha_1(T) \quad (\text{C.190})$$

$$\alpha_1(T - l(T) + 1) \leq \alpha_1(T + 1) \leq \alpha_1(T) \quad (\text{C.191})$$

$$\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1) \quad (\text{C.192})$$

Therefore, the two cases for which  $\alpha_1(T) \leq \alpha_1(T + 1)$  are:

- $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T))$ .
- $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1)$ .

Hence, according to the results of Proposition 5.13, the inequalities satisfied by  $\alpha_1(T + 1) - \alpha_1(T)$  are:

If  $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T))$ , then:

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T)) - \alpha_1(T)) \quad (\text{C.193})$$

If  $\alpha_1(T) \leq \alpha_1(T + 1) \leq \alpha_1(T - l(T) + 1)$ , then:

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(\alpha_1(T - l(T) + 1) - \alpha_1(T)) \quad (\text{C.194})$$

Since, by assumption of the Lemma,  $T \geq T_2 \geq T_\epsilon$ , then  $\max A_1(T) \leq l_2 + \epsilon$ . As a consequence,  $\alpha_1(T - l(T) + 1)$  and  $\alpha_1(T - l(T))$  which are elements of the vector  $A_1(T)$ , are less than  $l_2 + \epsilon$ .

Hence, for  $T \in [T_2, T_2 + L - 1]$ :

$$\alpha_1(T + 1) - \alpha_1(T) \leq p_1(l_2 + \epsilon - \alpha_1(T)) \quad (\text{C.195})$$

Therefore:

$$\alpha_1(T + 1) \leq p_1(l_2 + \epsilon) + (1 - p_1)\alpha_1(T) \quad (\text{C.196})$$

■

Now we should prove that for all possible sequences of  $\alpha_1$  in  $[T_2, T_2 + L]$ , their values can not exceed  $\lambda(1 - p_1)^L + (l_2 + \epsilon_2) = u_L$ .

**Lemma C.9.** *For all sequences of  $\alpha_1$  when  $T \in [T_2, T_2 + L]$ ,  $\alpha_1(T) \leq u_{T-T_2}$*

*Proof.* We prove this result by induction.

For  $T = T_2$ , we have that:

$$\alpha_1(T_2) \leq l_1 = u_0 \quad (\text{C.197})$$

We suppose that at time  $T$ ,  $\alpha_1(T) \leq u_{T-T_2}$ , then at time  $T + 1$ :

If  $\alpha_1(T + 1) \leq \alpha_1(T)$ :

Then as  $u_{T-T_2}$  is increasing in  $T$ :

$$\alpha_1(T + 1) \leq u_{T-T_2} \leq u_{T-T_2+1} \quad (\text{C.198})$$

If  $\alpha_1(T+1) \geq \alpha_1(T)$ :

Then, according to Lemma C.8:

$$\alpha_1(T+1) \leq p_1(l_2 + \epsilon) + (1 - p_1)\alpha_1(T) \quad (\text{C.199})$$

$$\leq p_1(l_2 + \epsilon) + (1 - p_1)u_{T-T_2} \quad (\text{C.200})$$

$$= u_{T-T_2+1} \quad (\text{C.201})$$

Therefore,  $\alpha_1(T+1) \leq u_{T-T_2+1}$ .

Hence, we have proved by induction that for all  $T \in [T_2, T_2 + L]$ ,  $\alpha_1(T) \leq u_{T-T_2}$  ■

As  $u_{T-T_2}$  is less than  $u_L$  for  $T \in [T_2, T_2 + L]$ , then according to Lemma C.9, the elements  $\alpha_1(T_2 + 1), \dots, \alpha_1(T_2 + L)$  are less than  $u_L < l_2$ .

Thus, we have found  $T_2 \geq T_\epsilon$  such that  $\alpha_1(T_2), \alpha_1(T_2 + 1), \dots, \alpha_1(T_2 + l_{\max})$  are strictly less than  $l_2$ . We denote  $T_2 + l_{\max}$  by  $T_d$  and we verify that  $\max A_1(T_d) < l_2$ . Indeed, we know that  $T_d - l(T_d) \geq T_d - l_{\max} = T_2$ , then the elements of the vector  $A_1(T_d)$  are included in the set of elements  $\{\alpha_1(T_2), \alpha_1(T_2 + 1), \dots, \alpha_1(T_2 + l_{\max})\}$ . That is  $\max A_1(T_d) < l_2$ . Hence, we have found  $T_d \geq T_\epsilon$ , such that  $\max A_1(T_d) < l_2$ .

## C.16 Proof of Proposition 5.15

In this proof, we show that for each state  $i$  in class  $k$ ,  $z_i^k(t)$  converges. To that end, we start first by specifying the eventual limit of  $z_i^k(t)$  for each  $i$ . To do so, we decompose  $1 - \alpha$  as follows:

$$l(p_1\alpha_1^* + p_2\alpha_2^*) + \gamma p_2\alpha_2^* + \beta p_1\alpha_1^* = 1 - \alpha \quad (\text{C.202})$$

where  $l$  is the biggest integer such that:  $l(p_1\alpha_1^* + p_2\alpha_2^*) < 1 - \alpha$ , and  $0 < \gamma \leq 1$  and  $\beta = 0$ ; or  $\gamma = 1$  and  $0 < \beta \leq 1$ . Then, we proceed with these following steps:

- We prove by induction that for all states  $1 \leq i \leq l + 1$ ,  $z_i^k(t)$  converges to  $p_k\alpha_k^*$ .
- Based on the theoretical findings of the first step, we prove that  $z_{l+2}^1(t)$  converges to  $(\beta + (1 - p_1)(1 - \beta))p_1\alpha_1^*$  and  $z_{l+2}^2(t)$  converges to  $(\gamma + (1 - p_2)(1 - \gamma))p_2\alpha_2^*$ .
- Finally, we show that for all states  $i > l + 2$ ,  $z_i^1(t)$  converges to  $(1 - p_1)^{i-l-2}(\beta + (1 - p_1)(1 - \beta))p_1\alpha_1^*$  and  $z_i^2(t)$  converges to  $(1 - p_2)^{i-l-2}(\gamma + (1 - p_2)(1 - \gamma))p_2\alpha_2^*$ .

1. For all states  $1 \leq i \leq l + 1$ ,  $z_i^k(t) \rightarrow p_k\alpha_k^*$ :

We prove this result by induction

- For  $i = 1$ , we have that  $z_1^k(t) = p_k\alpha_k(t - 1)$ . Therefore,  $z_1^k(t)$  converge to  $p_k\alpha_k^*$  as  $\alpha_k(t)$  converges to  $\alpha_k^*$ .
- We consider that for a certain  $j \leq l$ , for each  $1 \leq i \leq j$ ,  $z_i^k(t)$  converges to  $p_k\alpha_k^*$  and we show that  $z_{j+1}^k(t)$  converges also to  $p_k\alpha_k^*$ .  
Given that  $j \leq l$ :

$$j(p_1\alpha_1^* + p_2\alpha_2^*) < 1 - \alpha$$

We consider  $0 < \epsilon \leq 1 - \alpha - j(p_1\alpha_1^* + p_2\alpha_2^*)$ . Providing that  $z_i^k(t)$  converges to  $p_k\alpha_k^*$  for all  $1 \leq i \leq j$ , that means there exists  $t_j$  such that for  $t \geq t_j$ , for  $1 \leq i \leq j$ :

$$|z_i^k(t) - p_k\alpha_k^*| < \frac{\epsilon}{2j}$$

Hence:

$$\sum_{i=1}^j |z_i^1(t) - p_1\alpha_1^*| + \sum_{i=1}^j |z_i^2(t) - p_2\alpha_2^*| < \epsilon$$

That is,

$$\sum_{i=1}^j z_i^1(t) + \sum_{i=1}^j z_i^2(t) < \epsilon + j(p_1\alpha_1^* + p_2\alpha_2^*)$$

As consequence, for all  $t \geq t_j$ , we have that:

$$\sum_{i=1}^j z_i^1(t) + \sum_{i=1}^j z_i^2(t) < 1 - \alpha$$

Thus, for all  $t \geq t_j$ , the action prescribed to the users' proportion  $z_j^k(t)$  is the passive action <sup>1</sup>. Then, for all  $t \geq t_j$ :

$$z_{j+1}^k(t+1) = z_j^k(t)$$

Therefore,  $z_{j+1}^k(t)$  converges to  $p_k\alpha_k^*$ .

Consequently, we prove by induction that for all  $1 \leq i \leq l+1$ ,  $z_i^k(t)$  converges to  $p_k\alpha_k^*$ .

2.  $z_{l+2}^1(t) \rightarrow (\beta + (1 - p_1)(1 - \beta))p_1\alpha_1^*$  and  $z_{l+2}^2(t) \rightarrow (\gamma + (1 - p_2)(1 - \gamma))p_2\alpha_2^*$ .  
To avoid redundancy, we will be limited to the first case when  $0 < \gamma \leq 1$  and  $\beta = 0$ , since the proof's steps for both cases are exactly the same. We have that:

$$l(p_1\alpha_1^* + p_2\alpha_2^*) + \gamma p_2\alpha_2^* = 1 - \alpha$$

As  $\sum_{i=1}^l z_i^1(t) + \sum_{i=1}^l z_i^2(t)$  converges to  $l(p_1\alpha_1^* + p_2\alpha_2^*)$  which is strictly less than  $1 - \alpha$ , then there exists  $t_l$  such that for all  $t \geq t_l$ , we have that:

$$\sum_{i=1}^l z_i^1(t) + \sum_{i=1}^l z_i^2(t) < 1 - \alpha$$

As  $\sum_{i=1}^{l+1} z_i^1(t) + \sum_{i=1}^{l+1} z_i^2(t)$  converges to  $(l+1)(p_1\alpha_1^* + p_2\alpha_2^*)$  which is strictly greater than  $1 - \alpha$ , then there exists  $t_{l+1}$  such that for all  $t \geq t_{l+1}$ , we have that:

$$\sum_{i=1}^{l+1} z_i^1(t) + \sum_{i=1}^{l+1} z_i^2(t) > 1 - \alpha$$

---

<sup>1</sup>Knowing that the order of the proportions of the users according to the Whittle's index value alternates between the two classes in the set  $[1, l_{\max} + 1]$  as was established in 5.8, then for all integer  $b \in [1, l_{\max}]$ , the set  $\{z_i^k : k = 1, 2; 1 \leq i \leq b\}$  is the set of users with the lowest Whittle's index value. Therefore,  $\sum_{i=1}^b z_i^1(t) + \sum_{i=1}^b z_i^2(t) < 1 - \alpha$  implies that the actions prescribed to the users belonging to the set  $\{z_i^k : k = 1, 2; 1 \leq i \leq b\}$  is the passive action. By definition of  $l$ ,  $l < \frac{1-\alpha}{p_2\alpha_2^*}$ , then,  $l \leq l_{\max}$  (see Lemma 5.3). Hence, the above reasoning can be applied as well when  $b = l$ .

For  $t \geq \max\{t_l, t_{l+1}\}$ , we have that:

$$\sum_{i=1}^l z_i^1(t) + \sum_{i=1}^l z_i^2(t) < 1 - \alpha < \sum_{i=1}^{l+1} z_i^1(t) + \sum_{i=1}^{l+1} z_i^2(t)$$

Denoting  $\gamma(t)$  and  $\beta(t)$  the users' proportion of  $z_{l+1}^2(t)$  and  $z_{l+1}^1(t)$  respectively which are not scheduled, therefore, the relation that links  $z_{l+2}^1(t+1)$  and  $z_{l+2}^2(t+1)$  to  $z_{l+1}^1(t)$  and  $z_{l+1}^2(t)$  when  $t \geq \max\{t_l, t_{l+1}\}$ :

$$\begin{aligned} z_{l+2}^1(t+1) &= \beta(t)z_{l+1}^1(t) + (1-p_1)(1-\beta(t))z_{l+1}^2(t) \\ z_{l+2}^2(t+1) &= \gamma(t)z_{l+1}^2(t) + (1-p_2)(1-\gamma(t))z_{l+1}^1(t) \end{aligned}$$

with  $0 < \gamma(t) \leq 1$  and  $\beta(t) = 0$ ; or  $\gamma(t) = 1$  and  $0 < \beta(t) \leq 1$ . To that extent, we show that  $\beta(t)$  tends to  $\beta = 0$  and  $\gamma(t)$  tends to  $\gamma$ . For that purpose, we give the following equation which is always satisfied when  $t \geq \max\{t_l, t_{l+1}\}$ :

$$\sum_{i=1}^l z_i^1(t) + \sum_{i=1}^l z_i^2(t) + \gamma(t)z_{l+1}^2(t) + \beta(t)z_{l+1}^1(t) = 1 - \alpha \quad (\text{C.203})$$

Tending  $t$  to  $+\infty$  in the equation C.203, we obtain:

$$\lim_{t \rightarrow +\infty} [\gamma(t)z_{l+1}^2(t) + \beta(t)z_{l+1}^1(t)] = \gamma p_2 \alpha_2^*$$

We consider the set  $\{t : \beta(t) \neq 0\}$ . If this set is infinite, then there exists a strictly increasing function  $n(\cdot)$  from  $\mathbf{N}$  to  $\{t \in \mathbf{N} : \beta(t) \neq 0\}$ , such that  $\beta(n(t))$  is a subsequence of  $\beta(t)$ . As  $\beta(n(t)) \neq 0$ , then  $\gamma(n(t)) = 1$ . Therefore, we get:

$$\lim_{t \rightarrow +\infty} [z_{l+1}^2(n(t)) + \beta(n(t))z_{l+1}^1(n(t))] = \gamma p_2 \alpha_2^*$$

Since  $z_{l+1}^2(n(t))$  converges to  $p_2 \alpha_2^*$ , then:

$$\lim_{t \rightarrow +\infty} [\beta(n(t))z_{l+1}^1(n(t))] = (\gamma - 1)p_2 \alpha_2^*$$

$(\gamma - 1)p_2 \alpha_2^*$  is less than 0, and  $\beta(n(t))z_{l+1}^1(n(t))$  is greater than 0 for all  $t$ . Thus:

$$\lim_{t \rightarrow +\infty} [\beta(n(t))z_{l+1}^1(n(t))] = (\gamma - 1)p_2 \alpha_2^* = 0$$

This implies that  $\gamma = 1 = \gamma(n(t))$ , and  $\lim_{t \rightarrow +\infty} \beta(n(t)) = 0$  because  $z_{l+1}^1(n(t))$  converges to  $p_1 \alpha_1^* \neq 0$ . Hence  $\lim_{t \rightarrow +\infty} \beta(t) = 0 = \beta$ , i.e.  $\lim_{t \rightarrow +\infty} \gamma(t) = \gamma = 1$ .

If  $\{t : \beta(t) \neq 0\}$  is finite, then there exists  $t_e$  such that for all  $t \geq t_e$ ,  $\beta(t) = 0$ . Therefore, for all  $t \geq t_e$ , we have that:

$$\lim_{t \rightarrow +\infty} [\gamma(t)z_{l+1}^2(t)] = \gamma p_2 \alpha_2^*$$

That means  $\lim_{t \rightarrow +\infty} \beta(t) = 0$ , and  $\lim_{t \rightarrow +\infty} \gamma(t) = \gamma$ . Hence, in both cases,  $\beta(t) \rightarrow \beta = 0$  and  $\gamma(t) \rightarrow \gamma$ .

Consequently, combining the last result with the one derived in the first step, we conclude that  $z_{l+2}^1(t)$  converges to  $(\beta + (1-p_1)(1-\beta))p_1 \alpha_1^*$  and  $z_{l+2}^2(t)$  converges to  $(\gamma + (1-p_2)(1-\gamma))p_2 \alpha_2^*$ . Similar analysis can be applied to come with the aforementioned result when  $\gamma(t) = 1$  and  $0 < \beta(t) \leq 1$ .

3. For  $i > l + 2$ ,  $z_i^1(t) \rightarrow (1 - p_1)^{i-l-2}(\beta + (1 - p_1)(1 - \beta))p_1\alpha_1^*$  and  $z_i^2(t) \rightarrow (1 - p_2)^{i-l-2}(\gamma + (1 - p_2)(1 - \gamma))p_2\alpha_2^*$ :

For  $t \geq \max\{t_l, t_{l+1}\}$ , we are sure that the action prescribed to  $z_i^k(t)$  for all  $i \geq l + 2$  is the active action. As consequence,  $z_{i+1}^k(t + 1)$  satisfies:

$$z_{i+1}^k(t + 1) = (1 - p_k)z_i^k(t)$$

Therefore, as  $z_{l+2}^1(t)$  converges to  $(\beta + (1 - p_1)(1 - \beta))p_1\alpha_1^*$  and  $z_{l+2}^2(t)$  converges to  $(\gamma + (1 - p_2)(1 - \gamma))p_2\alpha_2^*$ , one can easily establish by induction that  $z_i^1(t)$  converges to  $(1 - p_1)^{i-l-2}(\beta + (1 - p_1)(1 - \beta))p_1\alpha_1^*$  and  $z_i^2(t)$  converges to  $(1 - p_2)^{i-l-2}(\gamma + (1 - p_2)(1 - \gamma))p_2\alpha_2^*$  for all  $i > l + 2$ .

We conclude that for all states  $i$  and  $k = 1, 2$ ,  $z_i^k(t)$  converges. On the other hands, according to Proposition 5.7, the only possible limit of  $\mathbf{z}(t)$  is  $\mathbf{z}^*$ . As consequence, for each  $k$  and  $i$ ,  $z_i^k(t)$  converges to  $z_i^{k,*}$ .

## C.17 Proof of Proposition 5.16

For a given  $\mathbf{z}$ , let  $m_1(\mathbf{z})$  and  $m_2(\mathbf{z})$  be the highest states of the class 1 and the class 2 respectively and  $l_1(\mathbf{z})$  and  $l_2(\mathbf{z})$  be the thresholds of class 1 and 2 respectively at time  $t$  when  $\mathbf{Z}^N(t) = \mathbf{z}$ . Given that, we introduce the following lemma.

**Lemma C.10.** *For any  $\mu$ , there exists positive constant  $C(\mathbf{z})$  such that:*

$$P(\|\mathbf{Z}^N(t + 1) - \mathbf{z}'\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{C(\mathbf{z})}{N} \quad (\text{C.204})$$

where  $C(\mathbf{z})$  is independent of  $N$  and  $\mathbf{z}' = Q(\mathbf{z})\mathbf{z} = \mathbf{E}(\mathbf{Z}^N(t + 1) | \mathbf{Z}^N(t) = \mathbf{z})$

*Proof.* By definition of  $m_1(\mathbf{z})$  and  $m_2(\mathbf{z})$ , we have that  $\mathbf{z} = (z_1^1, \dots, z_{m_1(\mathbf{z})}^1, z_1^2, \dots, z_{m_2(\mathbf{z})}^2)$ . On can easily show that  $m_1(\mathbf{z}') = m_1(\mathbf{z}) + 1$  and  $m_2(\mathbf{z}') = m_2(\mathbf{z}) + 1$  since the users' proportions at states  $m_1(\mathbf{z})$  and  $m_2(\mathbf{z})$  in class 1 and class 2 will become at states  $m_1(\mathbf{z}) + 1$  and  $m_2(\mathbf{z}) + 1$  at the next time slot respectively. To prove this lemma, we use the Chebychev inequality presented as follows:

$$P(|X - \mathbb{E}(X)| > \mu) \leq \frac{\text{Var}(X)}{\mu^2} \quad (\text{C.205})$$

for any  $\mu > 0$  and random variable  $X$ .

As  $\mathbf{z}' = \mathbf{E}(\mathbf{Z}^N(t + 1) | \mathbf{Z}^N(t) = \mathbf{z})$ , we can apply the Chebychev inequality. However we need to find the distribution of  $\mathbf{Z}^N(t + 1)$  knowing  $\mathbf{Z}^N(t) = \mathbf{z}$  in order to derive the expression of  $\text{Var}(\mathbf{Z}^N(t + 1) | \mathbf{Z}^N(t) = \mathbf{z})$ . It is more simple to study the parameters of one dimensional random variable than multi-dimensional random variable. Hence, instead of investigating  $\mathbf{Z}^N(t + 1)$ , we look into  $Z_i^{N,k}$ . In this regard, we have that:

$$\{\mathbf{Z}^N(t + 1) : \|\mathbf{Z}^N(t + 1) - \mathbf{z}'\| \geq \mu\} \subset \bigcup_{k,i} \{\mathbf{Z}^N(t + 1) : \|Z_i^{N,k}(t + 1) - z_i^{k,*}\| > \frac{\mu}{m_1(\mathbf{z}') + m_2(\mathbf{z}')}\} \quad (\text{C.206})$$

Therefore:

$$P(\|\mathbf{Z}^N(t+1) - \mathbf{z}'\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{z}) \leq P(\cup_{k,i} \{\|Z_i^{N,k}(t+1) - z_i'^k\| > \frac{\mu}{m_1(\mathbf{z}') + m_2(\mathbf{z}')}\} | \mathbf{Z}^N(t) = \mathbf{z}) \quad (\text{C.207})$$

$$\leq \sum_{k,i} P(\{\|Z_i^{N,k}(t+1) - z_i'^k\| > \frac{\mu}{m_1(\mathbf{z}') + m_2(\mathbf{z}')}\} | \mathbf{Z}^N(t) = \mathbf{z}) \quad (\text{C.208})$$

Now, we look for the distribution of  $Z_i^{N,k}(t+1)$  knowing  $\mathbf{Z}^N(t) = \mathbf{z}$ .

For  $2 \leq i \leq l_k(\mathbf{z})$ , as all the users at state  $i-1$  strictly less than  $l_k(\mathbf{z})$  will transit to the state  $i$  at the next time slot, then we have  $Z_i^{N,k}(t+1) = z_{i-1}^k = z_i'^k$ . This implies that:

$$P(\{\|Z_i^{N,k}(t+1) - z_i'^k\| > \frac{\mu}{m_1(\mathbf{z}') + m_2(\mathbf{z}')}\} | \mathbf{Z}^N(t) = \mathbf{z}) = 0 \quad (\text{C.209})$$

For  $i = 1$ , defining  $\alpha_1(\mathbf{z})$  and  $\alpha_2(\mathbf{z})$  as the proportions of the scheduled users in class 1 an class 2 respectively when  $\mathbf{Z}^N(t) = \mathbf{z}$ , then  $NZ_1^{N,k}(t+1) | \mathbf{Z}^N(t) = \mathbf{z}$  follows a binomial distribution with parameters  $p_k$  and  $\alpha_k(\mathbf{z})N$ . Therefore,  $Var(NZ_1^{N,k}(t+1) | \mathbf{Z}^N(t) = \mathbf{z}) = p_k(1-p_k)\alpha_k(\mathbf{z})N$ , which means that  $Var(Z_1^{N,k}(t+1) | \mathbf{Z}^N(t) = \mathbf{z}) = \frac{p_k(1-p_k)\alpha_k(\mathbf{z})}{N}$ . As a results, according to Chebychev inequality, we have that:

$$P(\{\|Z_1^{N,k}(t+1) - z_1'^k\| > \frac{\mu}{m_1(\mathbf{z}') + m_2(\mathbf{z}')}\} | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{p_k(1-p_k)\alpha_k(\mathbf{z})}{N\mu^2} (m_1(\mathbf{z}') + m_2(\mathbf{z}'))^2 \quad (\text{C.210})$$

For  $i \geq l_k(\mathbf{z}) + 2$ ,  $NZ_i^{N,k}(t+1) | \mathbf{Z}^N(t) = \mathbf{z}$  follows a binomial distribution with parameters  $1-p_k$  and  $z_{i-1}^k N$ . Hence,  $Var(Z_i^{N,k}(t+1) | \mathbf{Z}^N(t) = \mathbf{z}) = \frac{p_k(1-p_k)z_{i-1}^k}{N}$ . Thus:

$$P(\{\|Z_i^{N,k}(t+1) - z_i'^k\| > \frac{\mu}{i(m_1(\mathbf{z}') + m_2(\mathbf{z}'))}\} | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{p_k(1-p_k)z_{i-1}^k}{N\mu^2} (m_1(\mathbf{z}') + m_2(\mathbf{z}'))^2 i^2 \quad (\text{C.211})$$

Denoting  $\beta_k(\mathbf{z})$  the users' proportion of  $z_{l_k(\mathbf{z})}^k$  that will not be transmitted, then for  $i = l_k(\mathbf{z}) + 1$ ,  $NZ_i^{N,k}(t+1) | (\mathbf{Z}^N(t) = \mathbf{z}) = \beta_k(\mathbf{z})Nz_{i-1}^k + X$ , where  $X$  follows a binomial distribution with parameters  $1-p_k$  and  $(1-\beta_k(\mathbf{z}))z_{i-1}^k N$ , then:

$$P(\{\|Z_i^{N,k}(t+1) - z_i'^k\| > \frac{\mu}{i(m_1(\mathbf{z}') + m_2(\mathbf{z}'))}\} | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{p_k(1-p_k)(1-\beta_k(\mathbf{z}))z_{i-1}^k}{N\mu^2} (m_1(\mathbf{z}') + m_2(\mathbf{z}'))^2 i^2 \quad (\text{C.212})$$

We end up with:

$$\begin{aligned} & P(\|\mathbf{Z}^N(t+1) - \mathbf{z}'\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{z}) \\ & \leq (m_1(\mathbf{z}') + m_2(\mathbf{z}'))^2 \cdot \left[ \frac{p_1(1-p_1)\alpha_1(\mathbf{z})}{N\mu^2} + \frac{p_2(1-p_2)\alpha_2(\mathbf{z})}{N\mu^2} + \sum_{i \geq l_1(\mathbf{z})+2} \frac{p_1(1-p_1)i^2 z_{i-1}^1}{N\mu^2} \right. \\ & \left. + \sum_{i \geq l_2(\mathbf{z})+2} \frac{p_2(1-p_2)i^2 z_{i-1}^2}{N\mu^2} + \frac{p_1(1-p_1)(l_1(\mathbf{z})+1)^2(1-\beta_1(\mathbf{z}))z_{l_1(\mathbf{z})}^1}{N\mu^2} + \frac{p_1(1-p_2)(l_2(\mathbf{z})+1)^2(1-\beta_2(\mathbf{z}))z_{l_2(\mathbf{z})}^2}{N\mu^2} \right] \end{aligned} \quad (\text{C.213})$$

Knowing that  $\alpha_k(\mathbf{z}) \leq 1$ ,  $\sum_{i \geq l_k(\mathbf{z})} z_i^k \leq 1$ ,  $1 - \beta_k(\mathbf{z}) \leq 1$ , and for all state  $i$  in the vector  $\mathbf{z}'$ ,  $i \leq m_1(\mathbf{z}') + m_2(\mathbf{z}')$  then:

$$P(\|\mathbf{Z}^N(t+1) - \mathbf{z}'\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{(m_1(\mathbf{z}') + m_2(\mathbf{z}'))^4}{\mu^2 N} [2p_1(1-p_1) + 2p_2(1-p_2)]$$

Hence, denoting by  $C(\mathbf{z})$ ,  $\frac{(m_1(\mathbf{z}') + m_2(\mathbf{z}'))^4}{\mu^2} [2p_1(1-p_1) + 2p_2(1-p_2)] = \frac{(m_1(\mathbf{z}) + 1 + m_2(\mathbf{z}) + 1)^4}{\mu^2} [2p_1(1-p_1) + 2p_2(1-p_2)]$ , we obtain as a result:

$$P(\|\mathbf{Z}^N(t+1) - \mathbf{z}'\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{C(\mathbf{z})}{N} \quad (\text{C.214})$$

■

Now, we give a lemma that bounds the probability knowing the initial state  $\mathbf{z}(0) = \mathbf{x}$ . One can easily verify that  $m_1(\mathbf{z}(t)) = m_1(\mathbf{x}) + t$  and  $m_2(\mathbf{z}(t)) = m_2(\mathbf{x}) + t$  by induction. Without loss of generality, we let  $m_k(\mathbf{z}(t)) = m_k(t)$  for  $k = 1, 2$ .

**Lemma C.11.** *For any  $\mu$ , there exists positive constant  $C(t+1)$  such that:*

$$P_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu) \leq \frac{C(t+1)}{N} \quad (\text{C.215})$$

where  $C(t+1)$  is independent of  $N$ .

*Proof.* We recall from Lemma C.10 that for any  $\mu > 0$ , there exists a constant  $C(\mathbf{z})$  independent of  $N$  such that:

$$P(\|\mathbf{Z}^N(t+1) - Q(\mathbf{z})\mathbf{z}\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{C(\mathbf{z})}{N} \quad (\text{C.216})$$

Before proving the present lemma, we give an important lemma that will help us in the later analysis.

**Lemma C.12.** *For any proportion vector  $\mathbf{z}$ , there exists  $\sigma > 0$  such that if  $\|\mathbf{Z}^N(t) - \mathbf{z}\| \leq \sigma$ , then  $\mathbf{Q}(\mathbf{Z}^N(t)) = \mathbf{Q}(\mathbf{z})$ .*

*Proof.* One can deduce from the analysis done in Section 5.2, and more precisely the equation (5.26), that there exists  $\sigma > 0$  such that if  $\mathbf{Z}^N(t) \in \Omega_\sigma(\mathbf{z})$ ,  $\mathbf{Q}(\mathbf{Z}^N(t))$  is constant and doesn't depend on  $\mathbf{Z}^N(t)$ . Therefore, there exists  $\sigma > 0$  such that  $\mathbf{Q}(\mathbf{Z}^N(t)) = \mathbf{Q}(\mathbf{z})$ . That concludes the proof. ■

**Corollary C.1.** *For any  $v > 0$ , there exists  $\rho$  such that  $\|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \leq \rho \Rightarrow \|\mathbf{Q}(\mathbf{Z}^N(t))\mathbf{Z}^N(t) - \mathbf{Q}(\mathbf{z}(t))\mathbf{z}(t)\| \leq v$*

*Proof.* According to the previous lemma, if  $\|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \leq \sigma$ , then  $\mathbf{Q}(\mathbf{Z}^N(t)) = \mathbf{Q}(\mathbf{z}(t))$ . This implies that  $\|\mathbf{Q}(\mathbf{Z}^N(t))\mathbf{Z}^N(t) - \mathbf{Q}(\mathbf{z}(t))\mathbf{z}(t)\| = \|\mathbf{Q}(\mathbf{z}(t))\mathbf{Z}^N(t) - \mathbf{Q}(\mathbf{z}(t))\mathbf{z}(t)\| \leq \|\mathbf{Q}(\mathbf{z}(t))\| \|\mathbf{Z}^N(t) - \mathbf{z}(t)\|$ . That is, choosing  $\rho = \min\{\frac{v}{\|\mathbf{Q}(\mathbf{z}(t))\|}, \sigma\}$ , we get  $\|\mathbf{Q}(\mathbf{Z}^N(t))\mathbf{Z}^N(t) - \mathbf{Q}(\mathbf{z}(t))\mathbf{z}(t)\| \leq v$ . ■

With the above corollary being laid out, we prove the statement by a mathematical induction.

For  $t = 1$ , applying Lemma C.10, the following holds:

$$\begin{aligned} \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(1) - \mathbf{z}(1)\| \geq \mu) &= P(\|\mathbf{Z}^N(t+1) - Q(\mathbf{x})\mathbf{x}\| \geq \mu | \mathbf{Z}^N(t) = \mathbf{x}) \leq \frac{C(\mathbf{x})}{N} \\ &= \frac{C(1)}{N} \end{aligned} \quad (\text{C.217})$$

and the desired result holds for  $t = 1$  by simply choosing  $C(1) = \frac{(m_1(\mathbf{x})+1+m_2(\mathbf{x})+1)^4}{\mu^2} [2p_1(1-p_1) + 2p_2(1-p_2)]$ . Let us suppose that the statement holds for any  $t \geq 1$ . We investigate the property for  $t + 1$ . To that end, let us consider  $\nu < \mu$ . Therefore, according to Corollary C.1, there exists  $\rho$  such that:

$$\|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \leq \rho \Rightarrow \|\mathbf{Q}(\mathbf{Z}^N(t))\mathbf{Z}^N(t) - \mathbf{Q}(\mathbf{z}(t))\mathbf{z}(t)\| \leq \nu \quad (\text{C.218})$$

Bearing that in mind, we have that:

$$\begin{aligned} &\Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu) \\ &= \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \rho) \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \rho) \\ &\quad + \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \\ &\stackrel{(a)}{\leq} \frac{C'(t)}{N} + \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \end{aligned} \quad (\text{C.219})$$

where (a) follows from  $\Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \rho) \leq 1$  and  $C'(t)$  being the constant related to the statement holding for  $t$  and for  $\rho$ . Next, we tackle the second term of the inequality in (C.219):

$$\begin{aligned} &\Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \\ &= \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - Q(\mathbf{Z}^N(t))\mathbf{Z}^N(t) + Q(\mathbf{Z}^N(t))\mathbf{Z}^N(t) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \\ &\stackrel{(a)}{\leq} \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - Q(\mathbf{Z}^N(t))\mathbf{Z}^N(t)\| + \|Q(\mathbf{Z}^N(t))\mathbf{Z}^N(t) - Q(\mathbf{z}(t))\mathbf{z}(t)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \\ &\stackrel{(b)}{\leq} \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - Q(\mathbf{Z}^N(t))\mathbf{Z}^N(t)\| \geq \mu - \nu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \\ &= \sum_{\substack{\mathbf{z} \in \Omega_{\rho}(\mathbf{z}(t)) \\ m_k(\mathbf{z}) \leq m_k(\mathbf{z}(t)) \\ k=1,2}} \Pr_{\mathbf{x}}(\mathbf{Z}^N(t) = \mathbf{z} \mid \mathbf{Z}^N(t) \in \Omega_{\rho}(\mathbf{z}(t))) \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - Q(\mathbf{z})\mathbf{z}\| \geq \mu - \nu \mid \mathbf{Z}^N(t) = \mathbf{z}) \\ &\quad + \sum_{\substack{\mathbf{z} \in \Omega_{\rho}(\mathbf{z}(t)) \\ m_1(\mathbf{z}) > m_1(\mathbf{z}(t)) \\ \text{or} \\ m_2(\mathbf{z}) > m_2(\mathbf{z}(t))}} \Pr_{\mathbf{x}}(\mathbf{Z}^N(t) = \mathbf{z} \mid \mathbf{Z}^N(t) \in \Omega_{\rho}(\mathbf{z}(t))) \Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - Q(\mathbf{z})\mathbf{z}\| \geq \mu - \nu \mid \mathbf{Z}^N(t) = \mathbf{z}) \end{aligned} \quad (\text{C.220})$$

where (a) and (b) follows from the triangular inequality and the relationship in (C.218). One can notice that at any time slot  $t$ ,  $m_k(\mathbf{Z}^N(t)) \leq m_k(\mathbf{z}(t))$ . In light of that fact, the

second term of the equation (C.220) is equal to 0. Bearing that in mind, We have for  $\mathbf{z} \in \Omega_\rho(\mathbf{z}(t))$  such that  $m_k(\mathbf{z}) \leq m_k(\mathbf{z}(t))$ :

$$\Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - Q(\mathbf{z})\mathbf{z}\| \geq \mu - \nu | \mathbf{Z}^N(t) = \mathbf{z}) \leq \frac{C_1(\mathbf{z}(t))}{N} \quad (\text{C.221})$$

where  $C_1(t) = \frac{(m_1(\mathbf{z}(t))+m_2(\mathbf{z}(t))+2)^4}{(\mu-\nu)^2} [2p_1(1-p_1)+2p_2(1-p_2)] = \frac{(m_1(t)+m_2(t)+2)^4}{(\mu-\nu)^2} [2p_1(1-p_1)+2p_2(1-p_2)]$ . By substituting the above results in (C.220), we get:

$$\Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu \mid \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| < \rho) \leq \frac{C_1(t)}{N} \quad (\text{C.222})$$

Combining this with (C.219), we can conclude that there exists a constant  $C(t+1)$  such that:

$$\Pr_{\mathbf{x}}(\|\mathbf{Z}^N(t+1) - \mathbf{z}(t+1)\| \geq \mu) \leq \frac{C(t+1)}{N} \quad (\text{C.223})$$

which concludes our inductive proof. ■

Knowing that:

$$P_x\left(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu\right) \leq \sum_{t=0}^{T-1} P_x(\|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu)$$

Therefore, from Lemma C.11, there exists a constant  $C$  which doesn't depend on  $N$  such that:

$$P_x\left(\sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu\right) \leq \frac{C}{N}$$

Which concludes the proof.

## C.18 Proof of Lemma 5.4

We show first of all that  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$  with respect to our considered norm, i.e.  $\lim_{t \rightarrow +\infty} \sum_{i=1}^{+\infty} |z_i^k(t) - z_i^{k,*}|i = 0$  for  $k = 1, 2$ . For that purpose, we use the limit inversion theorem which states that:

- If the series  $\sum_i f_i(t)$  is uniformly convergent on  $\mathbb{R}^+$
- If for each integer  $i$ ,  $f_i(t)$  admits a finite limit  $r_i$  when  $t$  tends to  $+\infty$ .

Therefore,  $\lim_{t \rightarrow +\infty} \sum_{i=1}^{+\infty} f_i(t) = \sum_{i=1}^{+\infty} \lim_{t \rightarrow +\infty} f_i(t) = \sum_{i=1}^{+\infty} r_i$ .

By letting  $f_i(t)$  denotes  $|z_i^k(t) - z_i^{k,*}|i$  for a given  $k$ , proving the result above is equivalent to establish that:

$$\lim_{t \rightarrow +\infty} \sum_{i=1}^{+\infty} |z_i^k(t) - z_i^{k,*}|i = \sum_{i=1}^{+\infty} \lim_{t \rightarrow +\infty} |z_i^k(t) - z_i^{k,*}|i$$

To that extent, we check if the aforementioned conditions are satisfied for this specific function  $f_i(t) = |z_i^k(t) - z_i^{k,*}|i$ .

- Uniform convergence: According to Weierstrass criterion,  $\sum_i f_i(t)$  is uniformly convergent if for each  $i$  the function  $f_i(t)$  is bounded by a constant  $c_i$  such that  $\sum_i c_i$  is convergent. Based on the proof of the Proposition 5.15, one can deduce that for large enough  $t$  denoted by  $t_l$ , the following induction relation always holds for  $t \geq t_l$  and  $i \geq l_{\max} + 1$ :

$$z_{i+1}^k(t+1) = p_k z_i^k(t)$$

That is, choosing  $t_0$  greater than  $t_l$ , and denoting by  $i_0 = m_k(t_0)$  the highest state of the vector  $\mathbf{z}(t_0)$  which is greater than  $l_{\max} + 1$ , we have that for each  $i > i_0$ :

$$z_i^k(t) = \begin{cases} 0 & \text{if } t_0 \leq t < t_0 + i - i_0 \\ p_k^{i-i_0} z_{i_0}^k(t - (i - i_0)) & \text{if } t \geq t_0 + i - i_0 \end{cases} \quad (\text{C.224})$$

Based on the above equation, for each  $i > i_0$ ,  $z_i^k(t)$  is less than  $p_k^{i-i_0}$  for all  $t \geq t_0$ . To that extent, we investigate the evolution of the series of interest only when  $t \geq t_0$  (the limit inversion theorem still applicable since  $+\infty > t_0$ ). Moreover, we have that for all  $t \geq t_0$ :

$$\sum_i |z_i^k(t) - z_i^{k,*}|i = \sum_{i=1}^{i_0} |z_i^k(t) - z_i^{k,*}|i + \sum_{i=i_0+1}^{+\infty} |z_i^k(t) - z_i^{k,*}|i \leq i_0^2 + \sum_{i=i_0+1}^{+\infty} (p_k^{i-i_0} i + z_i^{k,*} i)$$

This last sum is known to be a finite sum since  $\sum_{i=1}^{+\infty} z_i^{k,*} i$  is the optimal average age of the relaxed problem for the class  $k$  which is finite, and  $\sum_{i=1}^{+\infty} p^i i$  is a finite sum for any  $0 \leq p < 1$ . Hence, the uniform convergence can be accordingly concluded.

- Existence of the limit of  $f_i(t) = |z_i^k(t) - z_i^{k,*}|i$ : According to the result of Proposition 5.15, we have  $\lim_{t \rightarrow +\infty} |z_i^k(t) - z_i^{k,*}|i = 0$  which is finite. Therefore, the second condition is satisfied.

Leveraging these findings, we can inverse the order between the limit and the sum. Subsequently:

$$\lim_{t \rightarrow +\infty} \sum_{i=1}^{+\infty} |z_i^k(t) - z_i^{k,*}|i = \sum_{i=1}^{+\infty} \lim_{t \rightarrow +\infty} |z_i^k(t) - z_i^{k,*}|i = 0$$

In other words, for  $k = 1, 2$ ,  $\sum_{i=1}^{+\infty} |z_i^k(t) - z_i^{k,*}|i$  tends to 0 when  $t$  grows. Consequently,  $\mathbf{z}(t)$  converges to  $\mathbf{z}^*$  with respect to our defined norm.

Therefore, for  $0 < \nu < \mu$ , there exists  $T_0$  such that for any  $t \geq T_0$ :

$$\|\mathbf{z}(t) - \mathbf{z}^*\| \leq \nu \quad (\text{C.225})$$

By leveraging Proposition 5.16, we have:

$$\begin{aligned} & \Pr_{\mathbf{x}} \left( \sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu \right) \\ & \leq \Pr_{\mathbf{x}} \left( \sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| + \|\mathbf{z}(t) - \mathbf{z}^*\| \geq \mu \right) \\ & \leq \Pr_{\mathbf{x}} \left( \sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu - \nu \right) \\ & \leq \Pr_{\mathbf{x}} \left( \sup_{0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}(t)\| \geq \mu - \nu \right) \leq \frac{s}{N} \end{aligned} \quad (\text{C.226})$$

which concludes the proof.

## C.19 Proof of Proposition 5.17

We have that:

$$\begin{aligned}
 & \left| \frac{1}{T} \mathbb{E}^{wi} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} Z_i^{k,N}(t) i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] - \sum_{k=1}^K \sum_{i=1}^{+\infty} z_i^{k,*} i \right| \\
 &= \left| \frac{1}{T} \mathbb{E}^{wi} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} (Z_i^{k,N}(t) i - z_i^{k,*} i) \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\
 &\leq \left| \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t) i - z_i^{k,*} i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\
 &\hspace{20em} \text{(C.227)}
 \end{aligned}$$

$$\begin{aligned}
 &+ \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t) i - z_i^{k,*} i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\
 &\hspace{20em} \text{(C.228)}
 \end{aligned}$$

We start by bounding (C.227). We have that:

$$\begin{aligned}
 & \left| \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t) i - z_i^{k,*} i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\
 &\leq \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ \left| Z_i^{k,N}(t) i - z_i^{k,*} i \right| \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] \\
 &\hspace{20em} \text{(C.229)}
 \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t) i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] + \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} z_i^{k,*} i \\
 &\hspace{20em} \text{(C.230)}
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{\max\{m_1(t), m_2(t)\}} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t) i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] + \frac{1}{T} \sum_{t=0}^{T_0-1} C^{RP} \\
 &\hspace{20em} \text{(C.231)}
 \end{aligned}$$

As  $m_k(\cdot)$  is increasing with  $t$ , then denoting  $m(t) = \max\{m_1(t), m_2(t)\}$ , we get:

$$\begin{aligned}
 & \frac{1}{T} \sum_{t=0}^{T_0-1} \sum_{k=1}^K \sum_{i=1}^{\max\{m_1(t), m_2(t)\}} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t) i \middle| \mathbf{Z}^N(0) = \mathbf{x} \right] + \frac{1}{T} \sum_{t=0}^{T_0-1} C^{RP} \leq \frac{(m(T_0) + C^{RP})T_0}{T} \\
 &\hspace{20em} \text{(C.232)}
 \end{aligned}$$

We denote  $Y_N$  the event  $\sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \geq \mu$ , and we proceed to bound the second

term (C.228).

$$\begin{aligned} & \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t)i - z_i^{k,*}i \mid \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\ &= P_{\mathbf{x}}(Y_N) \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t)i - z_i^{k,*}i \mid Y_N, \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \end{aligned} \quad (\text{C.233})$$

$$\begin{aligned} &+ (1 - P_{\mathbf{x}}(Y_N)) \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t)i - z_i^{k,*}i \mid \overline{Y}_N, \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\ & \hspace{15em} (\text{C.234}) \end{aligned}$$

$$\leq^{(a)} \frac{(T - T_0)(m(T) + C^{RP})}{T} P_{\mathbf{x}}(Y_N) + (1 - P_{\mathbf{x}}(Y_N))\mu \quad (\text{C.235})$$

where (a) results from:

$$\begin{aligned} & \left| \frac{1}{T} \sum_{t=T_0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} \mathbb{E}^{wi} \left[ Z_i^{k,N}(t)i - z_i^{k,*}i \mid \overline{Y}_N, \mathbf{Z}^N(0) = \mathbf{x} \right] \right| \\ & \leq \sup_{T_0 \leq t < T} \mathbb{E}^{wi} \left[ \sum_{k=1}^K \sum_{i=1}^{+\infty} |Z_i^{k,N}(t)i - z_i^{k,*}i| \mid \overline{Y}_N, \mathbf{Z}^N(0) = \mathbf{x} \right] \end{aligned} \quad (\text{C.236})$$

$$= \mathbb{E}^{wi} \left[ \sup_{T_0 \leq t < T} \|\mathbf{Z}^N(t) - \mathbf{z}^*\| \mid \overline{Y}_N, \mathbf{Z}^N(0) = \mathbf{x} \right] < \mu \quad (\text{C.237})$$

According to Lemma 5.4, we have  $\lim_{N \rightarrow \infty} P_{\mathbf{x}}(Y_N) = 0$ . Thus, combining the result (C.232) and (C.235), we obtain:

$$\lim_{N \rightarrow \infty} \left| \frac{1}{T} \mathbb{E}^{wi} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} Z_i^{k,N}(t)i \mid \mathbf{Z}^N(0) = \mathbf{x} \right] - \sum_{k=1}^K \sum_{i=1}^{+\infty} z_i^{k,*}i \right| \leq \frac{T_0(m(T_0) + C^{RP})}{T} + \mu \quad (\text{C.238})$$

This inequality is true for all  $\mu > 0$ , then:

$$\lim_{N \rightarrow \infty} \left| \frac{1}{T} \mathbb{E}^{wi} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} Z_i^{k,N}(t)i \mid \mathbf{Z}^N(0) = \mathbf{x} \right] - \sum_{k=1}^K \sum_{i=1}^{+\infty} z_i^{k,*}i \right| \leq \frac{T_0(m(T_0) + C^{RP})}{T} \quad (\text{C.239})$$

Finally we have:

$$\lim_{T \rightarrow \infty} \lim_{N \rightarrow \infty} \left| \frac{1}{T} \mathbb{E}^{wi} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} Z_i^{k,N}(t)i \mid \mathbf{Z}^N(0) = \mathbf{x} \right] - \sum_{k=1}^K \sum_{i=1}^{+\infty} z_i^{k,*}i \right| = 0 \quad (\text{C.240})$$

As consequence:

$$\lim_{T \rightarrow +\infty} \lim_{N \rightarrow \infty} \frac{1}{T} \mathbb{E}^{wi} \left[ \sum_{t=0}^{T-1} \sum_{k=1}^K \sum_{i=1}^{+\infty} Z_i^{k,N}(t)i \mid \mathbf{Z}^N(0) = \mathbf{x} \right] = \sum_{k=1}^K \sum_{i=1}^{+\infty} z_i^{k,*}i \quad (\text{C.241})$$