Asymptotically Optimal Delay-aware Scheduling in Queueing Systems

Saad Kriouile, Mohamad Assaad, Maialen Larranaga

Asymptotically Optimal Delay-aware Scheduling in Queueing Systems

Saad Kriouile[1], Mohamad Assaad [1], and Maialen Larranaga[2]

[1]Laboratoire des Signaux et Systèmes CentraleSupélec, 91192 Gif sur Yvette, France
[2]ASML, P.O. Box 324, 5500 AH Veldhoven, The Netherlands

***Abstract*—In this paper, we investigate a delay-aware channel allocation problem where the number of channels is less than that of users. Due to the proliferation of delay sensitive applications, the objective of our problem is chosen to be the minimization of the total average queuing delay of the network in question. First, we show that our problem falls 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. To that extent, we employ a Lagrangian relaxation for the original problem and prove it to be decomposable into multiple one-dimensional independent subproblems. Afterwards, we provide structural results on the optimal policy of each of the subproblems. More specifically, we prove that a threshold policy is able to achieve the optimal operating point of the considered subproblem. Armed with that, we show the indexability of the subproblems and characterize the Whittle's indices which are the basis of our proposed heuristic. We then provide a 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.**

I. Introduction

This paper deals with user and channel scheduling, which has been widely recognized as a means to improve the network performance and to meet the service demands of the users. This problem has been widely studied in the past and several allocation policies have been developed for various contexts (e.g. see [2]–[8] and the references therein). In 5G networks, the problem of channel and user scheduling will be receiving particular interest due to the increase in the number of devices and users. Furthermore, the applications nowadays do not need high data rates only but are also more delay-sensitive, which implies that minimizing the delay is considered as a main design metric in future networks.

In this paper, we consider the problem of scheduling and channel allocation in a discrete time system composed of one central scheduler serving multiple users or queues. We consider that the traffic arriving to each queue is time varying, and that the number of users is higher than the number of channels, which is a quite realistic assumption especially with the growth in density of users in today's networks. At each time slot, the central scheduler decides to allocate the channels to users, where a channel can be seen as a server in wired networks or a frequency bandwidth in wireless networks. Throughout this paper, we will use the terms "channel" and "server" interchangeably to designate a resource to allocate to users. Furthermore, we assume that the number of channels is limited and each channel can only allocate to one user at each time slot. The objective in this case is to find an allocation policy that minimizes the long-run average queuing delay of the users, as a mean to minimize the average delay in the network. Although it is a quite standard scheduling, we provide in this paper a rigorous mathematical analysis, leading to a novel scheduling algorithm of which we prove optimality in the many users regime. In fact, we show in this paper that the considered scheduling problem can be cast as a Restless Bandit Problem (RBP), which is a particular Markov Decision Process (MDP). However, RBPs are PSPACE-Hard (see Papadimitriou et al. [9]), and hence their optimal solution is out of reach. One should therefore propose suboptimal policies when dealing with such problems. In this paper, we approach the considered RBP problem using the Lagrangian relaxation technique, which consists of relaxing the constraint on the available resources. In other words, instead of having the constraint on the number of available channels satisfied in every time slot, we consider that it has to be satisfied on average. This allows us to decompose the large relaxed optimization problem into much simpler one-dimensional problems. Based on the optimal solution of the individual 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) and we will show that for our particular model, an explicit expression of the Whittle's index can be found. WIP has been proposed as a suboptimal policy for many problems in the literature, see for instance [10], [11]. It has also been shown to perform near optimally in many scenarios and in the particular case of multiclass M/M/1 queues, WIP which simplifies to the $c\mu$-rule is optimal, see Buyukkoc et al. [12], and Larranaga [13]. In this paper, we will prove that the developed WIP is asymptotically optimal in the many users regime. To that extent, we summarize in the following the key contributions of this paper:

- We provide an analysis of the relaxed optimization problem, which let us obtain the structure of the optimal solution of its dual problem. The optimal solution is shown to be a threshold-based policy by proving that the value function of the Bellman equation that resolves each individual dual problem satisfies the increasing property.
- We resolve the full balance equations verified by the stationary distribution of the Markov Chain representing the evolution of the queue state under a general threshold

policy n. This step is very crucial to derive the Whittle's indices expressions.
- We reformulate the individual dual problem of the relaxed problem using the steady state distribution. Afterwards, we provide a general algorithm that allows us to obtain the Whittle's index. To reduce even further the complexity, we provide a rigorous proof of the indexability of the classes, along with lemmas that allow us to derive simple expressions of the Whittle's index.
- We provide further characterization of the threshold-based optimal solution of the relaxed optimization problem. The structure of this solution helps us to prove the local asymptotic optimality of our proposed policy as we just need to compare the average cost under the Whittle's Index policy with the optimal cost of the relaxed problem. The reason behind that is the fact that the latter is always less than the optimal cost of the original problem.
- We show that the Whittle's Index policy is asymptotically optimal in the infinitely many users regime, that is, when the number of users in the system as well as the available channels grow large.
- Finally, we provide numerical performance results of the Whittle's Index policy that corroborate our claims.

A. Related Work

The problem of resource allocation and scheduling in wireless networks has been widely studied in the literature. In [2]–[6], throughput optimal schedulers have been derived for single channel, multi-channel and multi-user MIMO contexts. The aforementioned set of work focuses on developing strategies that stabilize the queues of the users using the max weight rule. The classical max weight rule is however known to be not delay optimal. To overcome this issue, many works have been developed in the past to take into account the average delay of the traffic of the users (e.g. see [14] and the references therein). Most of the existing works use Markov Decision Process (MDP) frameworks and develop allocation strategies using Bellman equation (e.g. by using value iteration, policy iteration, etc.). However, MDP frameworks and Bellman equation suffer from the curse of dimensionality, which leads to complex resource allocation strategies. In [15], [16], the authors try to minimize the average delay of the users' queues using Markov Decision Process (MDP) and stochastic learning tools. The complexity of the developed solutions is however much higher than the Whittle's index policy. Stochastic learning is also used in [17] to deal with the problem of power allocation in an OFDM (Orthgonal Frequency Division Multiplexing) system with the goal being to minimize the average delay of the users' packets in the queues. The developed solution requires high memory and computational complexity as compared to the Whittle's index policy.

On the other hands, [18]–[20] study a flow-Level scheduling problem with time-variant Markovian channels. They propose well-performing policies, but they don't tackle their optimality. For instance, in [18], the authors derive a well-performing policy called "Potential Improvement Rule" in the case where there is no arrivals and provide results regarding its optimality without giving analytical proofs and under several conditions specifically, when only one user can be served.

Whittle's index based policies have been used/developed in wireless networks to deal with the problem of pilot allocation over Markovian channel models. If a pilot is allocated to a user, its CSI can be estimated correctly and the user can hence transmit at a given rate. In [10], [21], a Gilbert-Elliot channel model is considered and the Whittle's index is derived. It has been shown in [21], [22] that a policy based on Whittle's index is asymptotically optimal for their specific problem. The authors in [23] extended the problem of pilot allocation to the case where the channel evolves according to a Markovian process between K states instead of two states as in the Gilbert-Elliot model. In the aforementioned papers, the queues of the users were not considered. In fact, the focus was on the channel allocation such that the long term total throughput (or equivalent objective function) is maximized without taking into account the dynamic traffic of the users. In this paper, the objective of the user/channel allocation is to minimize the long term average queuing delay of the users.

In [11], a derivation of the Whittle's 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's index policy has not been proved in [11] and the time was assumed to be continuous in their model. The authors in [24] considered the problem of project/job scheduling in which an effort is allocated to a fixed number of projects. The performance of a Whittle's index based policy was analyzed under a continuous time model. In contrast to these two papers, we consider that the time is slotted and that several users can be scheduled at a given time slot and not only one user. We provide an explicit characterization of the Whittle's indices, develop a Whittle's index channel allocation policy for our problem. On the other hand, in contrast to previous work in [1], in this paper, we consider that the buffer size is very tight. The motivation behind that is that in the framework of IoT (Internet of Things), the resources of the connected devices are very limited, especially the energy available and the backlog queue. To that extent, we suppose that the queue length of a given node does not exceed a certain constant denoted by L, which is in its turn less than the departure rate. Besides that, unlike our work in [1], we further prove the asymptotic optimality of our developed policy in the many user's regime.

The remainder of the paper is organized as follows: In Section II, we formulate the problem under investigation and we introduce the Lagrangian relaxation. In Section III, we prove the optimality of threshold/monotone policies for the relaxed problem. In Section IV, we compute the steady-state distribution of the system under a general threshold policy. In Section V, we characterize the Whittle's indices explicitly and we lay out our proposed Whittle's index based policy. Section VI provides further characterization of the optimal solution of the relaxed problem. In Sections VII and VIII, we prove the local and global asymptotic optimality of our proposed scheme respectively. In Section IX, we evaluate the performance of the Whittle's index policy numerically. Lastly,

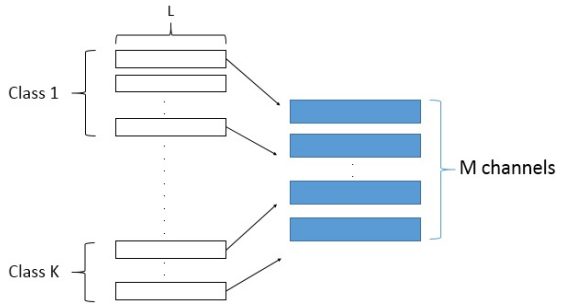

Figure 1: System Model

the mathematical proofs are provided in the appendices.

II. System Model and Problem Formulation

A. System model description

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 be used interchangeably throughout this paper, 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 for each user i in class-k, the number of arrival packet denoted by $A_i^k(t)$ follows a uniform distribution in $\{0, \dots, R_k - 1\}$ at each time slot t. Moreover, we consider that the buffer size of each user in the system is very tight. Accordingly, for each user in any class, if scheduled, transmits all the packets in its buffer. We denote by γ_k the proportion of class-k users in the system. We also let $q_i^{k,\phi}(t)$ denote the number of packets in queue i in class k. Furthermore, $s_i^{k,\phi}(\boldsymbol{q}^\phi(t))$ will denote the transmission action under a decision policy ϕ for user i in class k and $\boldsymbol{q}^\phi(t)$ 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 $s_i^{k,\phi}(t) := s_i^{k,\phi}(\boldsymbol{q}^\phi(t))$. If policy ϕ 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 less than R_k for all k. The general system model is presented in Figure 1. Based on our system model, the number of packets in queue i of class k evolves as follows:

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

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

B. Penalty function dynamics

In this paper, we are interested in minimizing the average delay incurred in the users' queues of the system. To that end, we should provide the expression of the queuing delay in function of the queue length. Then, we give the average cost function that we should minimize.

1) Queuing Delay metric: Given a user i, and class k, the average delay of each packet in this queue during the period $[0, T]$ is the delay's sum of all arrived packets between 0 and T which is, $\sum_{t=0}^{T} q_i^k(t)$ over the average number of the arrived packet between 0 and T. If there is no constraint on the buffer size as considered in [1], the average number of arrived packets between the time 0 and T is $\frac{T(R_k-1)}{2}$. Therefore, in this case, the number of arrived packets per time unit, or equivalently, the packet arrival rate is constant over time. To that extent, in our paper, where we consider that the buffer size is bounded by L, in order to have a fixed rate of the packet arrival as in [1], we take into consideration all the arrived packets, even those which are dropped due to the buffer constraint to give a simple expression of the delay metric. Since the surplus packets are immediately dropped, then we associate for this type of packets a delay that equals to 0. However, since the packets are dropped, which is an undesired event, a penalty should be incurred in the cost function that we will see later in the next section. As a consequence, the average delay according to the Little's Law is $\frac{2\sum_{t=0}^{T} q_i^k(t)}{T(R_k-1)}$. Denoting $\frac{2}{R_k-1}\beta_k$ by a_k where β_k is a weight factor, then the metric of interest is $\limsup_{T\to\infty} \frac{1}{T} \frac{\sum_{t=0}^{T} a_k q_i^k(t)}{T}$.

2) Penalty function: As mentioned in the section above, a penalty should be incurred when the packets are dropped. More precisely, a penalty should be paid when the number of arrived packets plus those in the queue exceeds L. To that extent, we should find for which state of the queue, the queue overflow will occur. Indeed, the only state that we can presume that there is an overflow at time t is when the queue state is equal to L. Bearing that in mind, we define for a given user i in class k, $b(q_i^k(t))$ that equals to 0 if $q_i^k(t) < L$ (zero penalty paid) and $b(l_i^k(t)) = C_d > 0$ (penalty is being paid). To that extent, we define the penalty function as follows:

$$d(q) = \begin{cases} q & if\ 0 \le q \le L-1 \\ L + C_d & if\ q = L \end{cases} \quad (2)$$

Consequently, the objective of the present work is to find a scheduling policy ϕ that minimizes the average weighted penalty function $\limsup_{T\to\infty} \frac{1}{T} \frac{\sum_{t=0}^{T} a_k d(q_i^k(t))}{T}$.

C. Problem formulation

The cost incurred by user i in class k, at time t is equal to $a_k d(q_i^{k,\phi}(t))$ for all $i \in \{1, \dots, \gamma_k N\}$. One can see that the model described in Section II belongs to the family of Restless Bandit Problems (RBP). We consider the broad class Φ 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 cost functions of different users, subject to the constraint on the number of users selected at each time slot. Given the initial state $\boldsymbol{q}(0) = (q_1^1(0), \dots, q_{N\gamma_1}^1(0), ..., q_1^K(0), \dots, q_{N\gamma_K}^K(0))$, the problem can be formulated as follows:

$$\min_{\phi\in\Phi} \limsup_{T\to\infty} \frac{1}{T} \mathbb{E}\left[\sum_{t=0}^{T-1} \sum_{k=1}^{K} \sum_{i=1}^{\gamma_k N} a_k d(q_i^{k,\phi}(t)) \mid \boldsymbol{q}(0)\right],$$
$$\text{s.t.} \sum_{k=1}^{K} \sum_{i=1}^{\gamma_k N} s_i^{k,\phi}(t) \le \alpha N, \text{ for all } t, \quad (3)$$

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

III. Relaxed Problem and Threshold-based Policy

As has been discussed in the introduction of this paper, RBPs are PSPACE-Hard (see Papadimitriou et al. [9]) and therefore one should develop well performing sub-optimal policies to solve these problems. In this paper, the development of our policy is done through several steps. First, we consider a Lagrangian relaxation of our problem and show that it can be decomposed into several one-dimensional problems. We then prove that the optimal solution to each of these relaxed problems is a threshold-based policy. We then compute the stationary distribution of the states of the system under the aforementioned threshold policy. This allows us to obtain a closed form expression of the Whittle's index values of the relaxed problem and develop a Whittle index-based scheduling policy for the original RBP.

In this section, we first formulate the relaxed problem and prove that its optimal policy is a threshold-based one.

A. 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 (3), has to be satisfied on average and not in every decision epoch, that is,

$$\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}s_i^{k,\phi}(t)\right]\leq\alpha N. \tag{4}$$

Note that, contrary to the strict constraint in Equation (3), the relaxed constraint allows the activation of more than α fraction of users at each time slot. If we note W the Lagrangian multiplier for the constrained problem, then the Lagrange function equals to:

$$f(W,\phi) \tag{5}$$

$$=\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}(a_k d(q_i^{k,\phi}(t))+W s_i^{k,\phi}(t))\mid\mathbf{q}_0\right] - W\alpha N, \tag{6}$$

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

$$\min_{\phi\in\Phi} f(W,\phi). \tag{7}$$

B. Problem Decomposition and Threshold-based Policy

In this section, we show that the relaxed problem can be decomposed into N one-dimensional subproblems, for which the optimal solution is a threshold-based policy. To do that, we first get rid of the constants that do not depend on ϕ and reformulate the problem as follows,

$$\min_{\phi\in\Phi}\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}(a_k d(q_i^{k,\phi}(t))+W s_i^{k,\phi}(t))\mid\mathbf{q}_0\right]. \tag{8}$$

One can see that the solution of this problem can be deduced from the well known Bellman equation (see Ross [25]). More specifically:

$$\bar{\boldsymbol{V}}(\mathbf{q})+\theta=\min_{\mathbf{s}}\{\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}C_k(q_i^k,s_i^k)+\sum_{\mathbf{q}'}Pr(\mathbf{q}'|\mathbf{q},\mathbf{s})\bar{\boldsymbol{V}}(\mathbf{q}')\}, \tag{9}$$

for all $\mathbf{q}=(q_1^1,\dots,q_{\gamma_1 N}^1,\dots,q_1^K,\dots,q_{\gamma_k N}^K)$, with $q_i^k\in\{0,\dots,L\}$ 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 (9), $\bar{\boldsymbol{V}}(\cdot)$ represents the Value Function, θ is the optimal average cost and $C_k(q_i^k,s_i^k)$ is the holding cost $a_k d(q_i^k)+W s_i^k$. The optimal decision for each state $\mathbf{q}$ can be obtained by minimizing the right hand side of Equation (9). We now show that the problem can be decomposed into N independent subproblems by decomposing $\bar{\boldsymbol{V}}(\cdot)$ into separate Value Functions for each user i in class k, i.e., $V_i^k(\cdot)$. In other words, the optimal decision $\mathbf{s}$ to problem (9) is a vector composed of elements s_i^k, where each s_i^k is nothing but the optimal decision that solves the individual Bellman equations.

$$V_i^k(q_i^k)+\theta_i^k=\min_{s_i^k}\{C_k(q_i^k,s_i^k)+\sum_{q_i'^k}Pr(q_i'^k|q_i^k,s_i^k)V_i^k(q_i'^k)\}. \tag{10}$$

Proposition 1. *Let $V_i^k(\cdot)$ be the optimal value function that solves Equation* (10)*, and let $\bar{\boldsymbol{V}}(\cdot)$ be the optimal value function that solves Equation* (9) *then:*

$$\bar{\boldsymbol{V}}(\cdot)=\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}V_i^k(\cdot). \tag{11}$$

Proof. See appendix A. □

In this section, we show that the solution to each individual 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. We first provide the definition of threshold policies.

Definition 1. *A threshold policy is a policy $\phi\in\Phi$ for which there exists $n\in\{-1,0,\cdots,L\}$ such that when the queue of user i is in state $q\leq n$, the prescribed action is $s^-\in\{0,1\}$, and when $q>n$, the prescribed action is $s^+\in\{0,1\}$ while bearing in mind that $s^-\neq s^+$.*
Since we only have two possible actions, a policy is of the form threshold policy if and only if it is monotone with q.

The solution of the Bellman equation (10) $V(\cdot)$ can be obtained by the well known Value iteration algorithm, which consists of updating $V_t(\cdot)$ using the following equation:

$$V_{t+1}(q)=\min_{s}\{C(q,s)+\sum_{q'}Pr(q'|q,s)V_t(q')\}-\theta \tag{12}$$

We consider that the initial value function V_0 is equal to 0 for any q, (i.e. for all q, $V_0(q)=0$). After many iterations, $V_t(\cdot)$ will converge to the unique fixed point of the equation (10) called $V(\cdot)$ (see [13, Chapter 1.3.3]). However, the value iteration algorithm is known to have high complexity and can take a long time to converge. Therefore, we will give

some structural properties of the value function $V_t(\cdot)$ for any t and conclude that the optimal policy is a threshold-based one.

Remark 1. *It is worth to emphasize that 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 queue length L, when the number of arrival packets A can be either $L-q$ or plus. Having said that, $Pr(L|q,1) = \sum_{j=L}^{R-1} Pr(A=j)$ and $Pr(L|q,0) = \sum_{j=L-q}^{R-1} Pr(A=j)$*

To establish our desired result, 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 [26]). We have by construction, $C(\cdot,s)$ increases with q. Since θ is just a constant, $V_{t+1}(\cdot)$ will be as well an increasing function with q.
As a 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 the next state before the arrival of the packets will be $q=0$ if the action prescribed is the active action since all the packets 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 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:

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

Bearing in mind that $V(\cdot)$ is increasing with q, then again according to Puterman [26], $\sum_{q'} Pr(q'|q,0)V(q')$ 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 (9), is the one that minimizes $V^s(q)$. Explicitly, $s(q) = \text{argmin}\{V^0(q), V^1(q)\}$. Moreover, exploiting the fact that $V^1(q)-V^0(q)$ is decreasing with q, then there exists $q_0 \in \{-1,\cdots,L\}$ 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 (8) is of type threshold increasing policy.

IV. Stationary distribution

We have seen previously that the optimal solution of Problem (8) is a threshold-based policy. Let us define n_k as the threshold for users in class k, i.e. if the queue state of user i in class k is q_i^k such that $q_i^k \leq n_k$ 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. This will be useful in the subsequent section in the derivation of a closed form expression of the Whittle's index values. We assume here that at each queue i in class k, packets arrive according to a discrete uniform distribution, that is, $\mathbb{P}(A_i^k(t)=x)=\rho_k$ for all $0\leq x\leq R_k-1$ and 0 otherwise, where $\rho_k = 1/R_k$. For ease of notation, we again drop the indices k and i (e.g. we denote the threshold by n and the queue state by q). To that extent, we denote by $p_n(i,j)$ the transition probability from queue state i to queue state j, by $u(.)$ the stationary distribution under the threshold policy n, and by $R-1$ the maximum arrival rate ($\rho = 1/R$). Finding the stationary distribution requires resolving the full balance equation:

$$u(i) = \sum_{j=0}^{n} p_n(j,i)u(j) \tag{13}$$

In most of works in literature, the authors consider that the evolution of the metric under the passive action is deterministic [21], [23], [27], [28]. In other words, for each state i, we know for sure the next state to which the bandit will transit under the passive action (usually $i+1$). Explicitly, $p_n(j,j+1)=1$ if $j\leq n$. That explains why in these papers above, the authors got a simple recurrence relation between the elements of the stationary distribution under a given threshold policy. For instance, in [27], leveraging only the evolution of AoI (Age of information metric), the authors got directly $u(i+1)$ in function of $u(i)$ without requiring any further manipulations. Whereas in this paper, obtaining a recurrence relation is not straightforward, since the expression of $u(i)$ is a linear combination of $\{u(j)\}_{j\in A_i}$ where the cardinal of the set A_i is at least $L-n$ as we will see later.

$u(.)$ verifies the following full balance equation:

$$u(i) = \sum_{j=0}^{L} p_n(j,i)u(j) = \sum_{j=0}^{n} p_n(j,i)u(j) + \sum_{j=n+1}^{L} p_n(j,i)u(j) \tag{14}$$

Definition 2. *We define π_i as:*

$$\pi_i = \begin{cases} \rho & if\ 0\leq i\leq R-1 \\ 0 & else \end{cases} \tag{15}$$

Proposition 2. *The expressions of $p_n(j,i)$ are given by:*
If $0 \leq i < L$ and $j \leq n$

$$p_n(j,i) = \pi_{i-j} = \begin{cases} \rho & if\ 0 \leq i-j \leq R-1 \\ 0 & else \end{cases} \tag{16}$$

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

$$p_n(j,i) = \pi_i = \begin{cases} \rho & if\ 0 \leq i \leq R-1 \\ 0 & else \end{cases} \tag{17}$$

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

$$p_n(j,L) = (R-L+j)\pi_{L-j} = (R-L+j)\rho \tag{18}$$

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

$$p_n(j,L) = (R-L)\pi_L = (R-L)\rho \tag{19}$$

Proof. See appendix B. □

Proposition 3. *The expressions of the stationary distribution is:*
1) $-1 \leq n \leq L-1$:

$$u(i) = \begin{cases} \rho(1-\rho)^{n-i} & if\ 0 \leq i \leq n \\ \rho & if\ n+1 \leq i \leq L-1 \\ (1-\rho)^{n+1} - (L-n-1)\rho & if\ i = L \end{cases} \tag{20}$$

2) $n = L$:

$$u(i) = \begin{cases} 0 & if\ 0 \leq i \leq L-1 \\ 1 & if\ i = L \end{cases} \tag{21}$$

Proof. See appendix C. □

V. Whittle's index

In this section, we provide the derivation of the Whittle's indices, which are values that depend on the queue state of the user. 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 individual dual relaxed problem at this state is indifferent (passive and active decision are both optimal). However, the Whittle's index is well defined only if the property of indexability is satisfied. This property requires to establish 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 work on a given class k, and we consider its maximum arrival rate is $R-1$ ($a = \frac{2}{R-1}$) with $\rho = 1/R$. All the obtained results here can be applied for any class.

We start the derivation by first reformulating the dual of the relaxed problem using the stationary distribution derived in the previous section. Since the solution of the dual of the relaxed problem (8) (given a constant W) is a threshold-based policy, we can reformulate the problem as follows:

$$\min_{n\in[0,L]} \mathbb{E}[a\ d(q^n) + W s^n] \tag{22}$$

$$= \min_{n\in[0,L]} \{\sum_{i=0}^{L} a u^n(i) d(i)\ -\ W \sum_{q=0}^{n} u^n(i) + W\} \tag{23}$$

with n and u^n being the threshold and the stationary distribution under the threshold policy n with respect to the queue length respectively.
The new formulation of the problem turns out to be useful to derive the Whittle's indices since, for any W, we can find the minimizer of the expression in equation (22).
We first give the expression of the mean cost in equation (22) given threshold policy n (for all possible values of n and L). If $-1 \leq n \leq L-1$:

$$\sum_{i=0}^{L} a u^n(i) d(i) = a[(L+R+C_d)(1-\rho)^{n+1} + n - R + 1 - (L-n-1)\rho C_d + \frac{(L-1-n)(n-L)}{2R}] \tag{24}$$

If $n = L$:

$$\sum_{i=0}^{L} a u^n(i) d(i) = aL + aC_d \tag{25}$$

Second, we provide the expression of the passive decision's average time in equation (22) given a threshold n:

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

$$\sum_{i=0}^{n} u^n(i) = 1 - (1-\rho)^{n+1} \tag{26}$$

If $n = L$:

$$\sum_{i=0}^{n} u^n(i) = 1 \tag{27}$$

A. Computation of the Whittle's index values

We first formalize the indexability and the Whittle's index in the following definitions.

Definition 3. *Considering problem* (22) *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* (22)*) 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 (22) is a stationary policy, more precisely, a threshold based policy.

Definition 4. *A class is indexable if the set of states in which the passive action is the optimal action increases with 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:*

$$W(n) = \min\{W | n \in D(W)\} \tag{28}$$

In the literature, several works have been conducted to find the Whittle's index values. For example, an interesting iterative algorithm has been provided in [13]. Even though the context of our work here is different from the one considered in [13], we will prove in the sequel that the proposed algorithm in [13] can be adapted to our case up to some modifications (in our case we have a maximum buffer state L). In addition, further analysis will be provided here to derive a closed form expression of the Whittle's index values. We will first provide this modified algorithm and then prove that it allows the computation of the Whittle's index values for our problem.

Algorithm 1 Whittle Index Computation

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

Proposition 4. *Assuming that the optimal solution is a threshold policy, and that $\sum_{q=0}^{n} u^n(q)$ is increasing, then the class is indexable. Moreover, if $\sum_{q=0}^{L} au^n(q)d(q)$ is increasing with n and for all i and j such that $i < j$ $\sum_{q=0}^{i} u^i(q) = \sum_{q=0}^{j} u^j(q) \implies \sum_{q=0}^{L} au^i(q)d(q) < \sum_{q=0}^{L} au^j(q)d(q)$, then the Whittle's index values are computed by applying Algorithm 1.*

Proof. For the proof, see appendix D. □

Remark 2. *In order to simplify the notation in the sequel, we denote $\sum_{q=0}^{L} au^n(q)d(q)$ by a_n and $\sum_{q=0}^{n} u^n(q)$ by b_n.*

In order to apply Algorithm 1 that allows to obtain the Whittle's index for each state in our case, we need to prove that the conditions given in Proposition 4 are satisfied.

We start first by establishing the indexability.

Theorem 1. *For each k, the class-k is indexable.*

Proof. According to Proposition 4, we just need to prove that $\sum_{q=0}^{n} u^n(q)$ is increasing with n. It is clear that from Equation (26), $\sum_{q=0}^{n} u^n(q)$ is increasing with n. Hence, the class is indexable. □

We prove the two others conditions of Proposition 4 which are the increase property of $\sum_{q=0}^{L} au^n(q)d(q)$ with n, and that for all i and j such that $i < j$, $\sum_{q=0}^{i} u^i(q) = \sum_{q=0}^{j} u^j(q) \implies \sum_{q=0}^{L} au^i(q)d(q) < \sum_{q=0}^{L} au^j(q)d(q)$. 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 5. *$a_n = \sum_{q=0}^{L} au^n(q)d(q)$ is increasing with n.*

Proof. See appendix G. □

As the indexability is satisfied and the two conditions of Proposition 4 are verified, then we can apply Algorithm 1 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's indices.
We first proceed by laying out the following definitions and lemmas.

Definition 5. *For any given increasing threshold policy n, we define y^n as a function of the subsidy W, such that $y^n(W) = \sum_{q=0}^{L} au^n(q)d(q) - W\sum_{q=0}^{n} u^n(q) = a_n - Wb_n$.*

Lemma 1. *The intersection point $W = x_{i,j}$ between $y^i(W)$ and $y^j(W)$ is equal to:*

$$x_{i,j} = \frac{\sum_{q=0}^{L} au^i(q)d(q) - \sum_{q=0}^{L} au^j(q)d(q)}{\sum_{q=0}^{i} u^i(q) - \sum_{q=0}^{j} u^j(q)} \tag{29}$$

Proof. See Appendix H □

Theorem 2. *The Whittle's index of state $n \in [0, L]$:*

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

Proof. See appendix I □

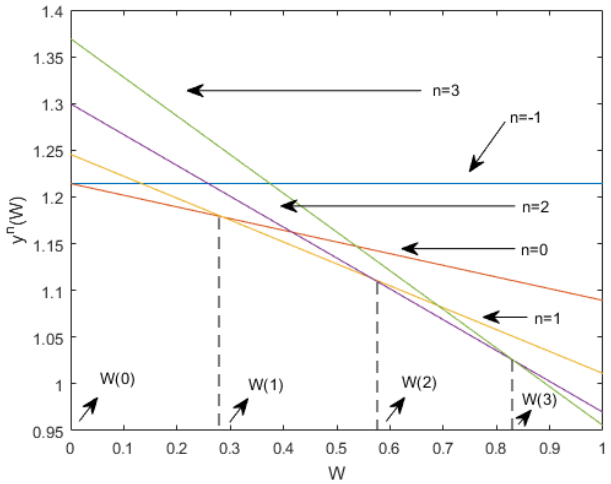

Figure 2: $y^n(W)$ in function of W for different values of n ($L = 4$, $R - 1 = 7$ and $C_d = 3$): $W(i)$ indicates the Whittle's index at state i

B. Whittle's index policy for the original problem

We now consider the original optimization problem (3) and propose a simple Whittle's index policy. This policy consists of simply allocating the channels to the M users with the highest Whittle's indices at time t, denoted by WIP, and computed using the simple expressions in Theorem 2.

VI. Further analysis of the optimal solution of the relaxed problem

In this section, we provide further analysis and give the structure of the optimal solution for the relaxed problem, which will be useful for the proof of optimality of the Whittle's Index policy. As we have seen in Section III, for any given W, the optimal solution for the dual relaxed problem (8) is a threshold-based policy for each user. By using the Whittle's index expressions derived in Theorem 2, 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's index at state i in class k. We denote by $l = (l_1, l_2, \cdots, l_K)$ 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 6. *For a given W, the optimal threshold vector $l = (l_1(W), l_2(W), \cdots, l_K(W))$ for the dual problem satisfies:*
For each k:

$$l_k(W) = \arg\max_i \{W_i^k | W_i^k \leq W\} \tag{30}$$

or

$$l_k(W) = \arg\max_i \{W_i^k | W_i^k < W\} \tag{31}$$

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 J. □

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

Proposition 7. *The solution of the constrained relaxed problem is of type threshold policy $l(W^*)$, with l being the function vector defined in Proposition 6 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 K. □

However, W^* that satisfies the above constraint may not exist since α is a real number that can take any value in $[0, 1]$, and $\sum_{k=1}^{K} \gamma_k \sum_{i=l_{k+1}(W)}^{L} u_k^{l_k(W)}(i)$ is discrete, since the vector $l(W)$ can only take discrete values in $[0, L]^K$. 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 has been mentioned in Proposition 6. To that extent, our task is to find among these values of W, the one for which there exists a randomized parameter θ such that the constraint is satisfied with equality. To that end, we introduce this following proposition.

Proposition 8. *There exists a class m, state p, and a randomization parameter θ^* 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 θ^* 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) *is satisfied with equality, i.e.*

$$\alpha = \sum_{k \neq m} \sum_{i=l_k(W_p^m)+1}^{L} \gamma_k u_k^{l_k(W_p^m)}(i) + \sum_{i=l_m(W_p^m)+1}^{L} \gamma_m u_m^*(i) + (1-\theta^*)\gamma_m u_m^{l_m(W_p^m)-1}(l_m(W_p^m)) \tag{32}$$

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

Proof. See appendix L □

The solution of the dual problem described in Proposition 8 satisfies the constraint (4) with equality, then according to Proposition 7, 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:

$$C^{RP,N} = \sum_{k \neq m} \sum_{i=0}^{L} N\gamma_k a_k u_k^{l_k(W_p^m)}(i) d(i) + \sum_{i=0}^{L} N\gamma_m a_m u_m^*(i) d(i) \tag{33}$$

VII. Local optimality

In this section, we will show that the performance of the Whittle's Index policy is asymptotically locally optimal. The asymptotic optimality means that for a large number of users N and a large number of channels M ($\alpha = \frac{M}{N}$ is a constant value), the Whittle's Index policy is optimal. For that we will compare the average cost obtained by the Whittle's Index policy WIP with the one obtained for the relaxed problem RP. 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. The reason behind comparing $C^{RP,N}$ and $C_T^N(\mathbf{x})$ is that $C^{RP,N}$ is a lower bound of all expected average cost obtained by any policy that resolves the original Problem (3). This means that it is sufficient to prove that $C_T^N(\mathbf{x})$ converges to $C^{RP,N}$ when T and N scale in order to establish the asymptotic optimality of Whittle's Index policy. For that, we will be in need of

the optimal cost expression of the relaxed problem $C^{RP,N}$ derived in Section VI.

First, we denote by $Z_i^{k,N}$ the proportion of queues at state i in class k over all the queues of the system. In other words, it denotes the number of queues at state i in class k over the number of all users which is N. We have that $\mathbf{Z}^N = (\mathbf{Z}^{1,N}, ..., \mathbf{Z}^{K,N})$ with $\mathbf{Z}^{k,N} = (Z_1^{k,N}, ..., Z_L^{k,N})$ 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 $\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{L} a_k Z_i^{k,N}(t)d(i)N \mid \mathbf{Z}^N(0) = \mathbf{x}\right]$, where $\mathbf{Z}^N(t)$ evolves under Whittle's Index policy. Denoting by $\mathbf{z}^*$ the optimal proportion of the 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 that consists of analyzing the evolution of the expectation of $\mathbf{Z}^N(t)$ under the Whittle's Index policy. For that, we define the vector $\mathbf{z}(t)$ as follows:

$$\boldsymbol{z}(t+1) - \boldsymbol{z}(t)|_{\boldsymbol{z}(t)=\boldsymbol{z}} = \mathbb{E}\left[\boldsymbol{Z}^N(t+1) - \boldsymbol{Z}^N(t) | \boldsymbol{Z}^N(t) = \boldsymbol{z}\right] \tag{34}$$

If we denote by w_j^h the Whittle's index for class h at state j and by $p_i^k(\boldsymbol{z})$ the probability that a user is selected randomly among z_i^k to transmit, one can easily show that [24]:

$$p_i^k(\boldsymbol{z}) = \min\{z_i^k, \max(0, \alpha - \sum_{w_j^h > w_i^k} z_j^h)\}/z_i^k \tag{35}$$

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:

$$q_{i,j}^k(\boldsymbol{z}) = p_i^k(\boldsymbol{z}) q_{i,j}^{k,1} + (1 - p_i^k(\boldsymbol{z})) q_{i,j}^{k,0} \tag{36}$$

Accordingly, we have that for each i and k:

$$z_i^k(t+1) - z_i^k(t) = \sum_{j \neq i} q_{j,i}^k(\boldsymbol{z}(t)) z_j^k(t) - \sum_{i \neq j} q_{i,j}^k(\boldsymbol{z}(t)) z_i^k(t) \tag{37}$$

Let w^* be the Lagrangian parameter that gives the optimal solution of the relaxed problem. Then, according to Proposition 8, 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 by l_k, and a randomized policy between two threshold policies l_m and $l_m - 1$ for class m.

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

Providing that for all k and t:

$$\sum_{j=0}^{L} z_j^k(t) = \gamma_k \tag{38}$$

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

1) $k \neq m$:

$$z_i^k(t+1) \tag{39}$$
$$= \sum_{j=0}^{l_k - 1} (q_{j,i}^{k,0} - q_{l_k,i}^{k,0}) z_j^k(t) + \sum_{j=l_k+1}^{L} (q_{j,i}^{k,1} - q_{l_k,i}^{k,0}) z_j^k(t) \tag{40}$$
$$+ \gamma_k q_{l_k,i}^{k,0} \tag{41}$$

2) $k = m$:

$$\begin{aligned} & z_i^m(t+1) \\ & = \sum_{j=0}^{l_m - 1} (q_{j,i}^{m,0} - q_{l_m,i}^{m,0}) z_j^m(t) + \sum_{j=l_m+1}^{L} (q_{j,i}^{m,1} - q_{l_m,i}^{m,1}) z_j^m(t) \\ & + (1-\alpha) q_{l_m,i}^{m,0} + \alpha q_{l_m,i}^{m,1} - \Big(\sum_{\substack{w_j^h > w_{l_m}^m \\ h \neq m, j \neq l_h}} z_j^h(t) \Big) q_{l_m,i}^{m,1} \\ & - \Big(\sum_{\substack{w_j^h \leq w_{l_m}^m \\ h \neq m, j \neq l_h}} z_j^h(t) \Big) q_{l_m,i}^{m,0} + \Big(\sum_{\substack{h=1 \\ h \neq m}}^{K} \sum_{\substack{j=0 \\ j \neq l_h}}^{L} \mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} z_j^h(t) \Big) q_{l_m,i}^{m,1} \\ & + \Big(\sum_{\substack{h=1 \\ h \neq m}}^{K} \sum_{\substack{j=0 \\ j \neq l_h}}^{L} \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} z_j^h(t) \Big) q_{l_m,i}^{m,0} \\ & - \sum_{\substack{h=1 \\ h \neq m}}^{K} \gamma_h (\mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} q_{l_m,i}^{m,1} + \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} q_{l_m,i}^{m,0}) \end{aligned} \tag{42}$$

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 $C = (c^1, \cdots, c^K)$ such that $c^k = (\gamma_k q_{l_k,0}^{k,0}, \cdots, \gamma_k q_{l_k,L}^{k,0})$ and $c^m = ((1-\alpha) q_{l_m,0}^{m,0} + \alpha q_{l_m,0}^{m,1} - g_0^m, \cdots, (1-\alpha) q_{l_m,L}^{m,0} + \alpha q_{l_m,L}^{m,1} - g_L^m)$ for each $k \neq m$.

Then:

1) $k \neq m$:

$$z_i^k(t+1) = \sum_{j=0}^{l_k - 1} (q_{j,i}^{k,0} - q_{l_k,i}^{k,0}) z_j^k(t) + \sum_{j=l_k+1}^{L} (q_{j,i}^{k,1} - q_{l_k,i}^{k,0}) z_j^k(t) + c_i^k \tag{43}$$

2) $k = m$:

$$
\begin{aligned}
&z_i^m(t+1)\\
&= \sum_{j=0}^{l_m-1}(q_{j,i}^{m,0} - q_{l_m,i}^{m,0})z_j^m(t) + \sum_{j=l_m+1}^{L}(q_{j,i}^{m,1} - q_{l_m,i}^{m,1})z_j^m(t) + c_i^m\\
&- \Big(\sum_{\substack{w_j^h > w_{l_m}^m\\ h\neq m, j\neq l_h}} z_j^h(t)\Big)q_{l_m,i}^{m,1} - \Big(\sum_{\substack{w_j^h \leq w_{l_m}^m\\ h\neq m, j\neq l_h}} z_j^h(t)\Big)q_{l_m,i}^{m,0}\\
&+ \Big(\sum_{\substack{h=1\\ h\neq m}}^{K}\sum_{\substack{j=0\\ j\neq l_h}}^{L} \mathbb{1}_{\{w_{l_h}^h > w_{l_m}^m\}} z_j^h(t)\Big)q_{l_m,i}^{m,1}\\
&+ \Big(\sum_{\substack{h=1\\ h\neq m}}^{K}\sum_{\substack{j=0\\ j\neq l_h}}^{L} \mathbb{1}_{\{w_{l_h}^h \leq w_{l_m}^m\}} z_j^h(t)\Big)q_{l_m,i}^{m,0} \qquad (44)
\end{aligned}
$$

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 $\jmath_{w^*}$ between $\tilde{\boldsymbol{z}}(t+1)$ and $\tilde{\boldsymbol{z}}(t)$ where $\tilde{\boldsymbol{z}}$ is the proportion vector in which the elements $z_{l_k}^k$ for different k are eliminated.

$$\tilde{\boldsymbol{z}}(t+1) = \boldsymbol{Q}\tilde{\boldsymbol{z}}(t) + \boldsymbol{C} \qquad (45)$$

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

$$\tilde{\boldsymbol{z}}(t) - \tilde{\boldsymbol{z}}^* = \boldsymbol{Q}^t\boldsymbol{e} \qquad (46)$$

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

Lemma 2. *If for all eigenvalues λ of $\boldsymbol{Q}$, $|\lambda| < 1$, then there exists a neighborhood $\Omega_\sigma(\tilde{\boldsymbol{z}}^*) \subseteq \jmath_{w^*}$ such that if $\tilde{\boldsymbol{z}}(0) \in \Omega_\sigma(\tilde{\boldsymbol{z}}^*)$, we have the following:*
1) For all $t \geq 0$, $||\tilde{\boldsymbol{z}}(t) - \tilde{\boldsymbol{z}}^*|| < \sigma$ ($\tilde{\boldsymbol{z}}(t) \in \jmath_{w^*}$).
2) $\tilde{\boldsymbol{z}}(t)$ converges to $\tilde{\boldsymbol{z}}^*$.

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

Proposition 9. *For all eigenvalue λ of $\boldsymbol{Q}$, $|\lambda| < 1$*

Proof. See the proof in appendix M. □

The aforementioned result, combined with Lemma 2, proves the convergence of the fluid limit system. Consequently, $\boldsymbol{z}(t)$ converges to the fixed point of Equation (34), $\boldsymbol{z}^*$. However, the above result is not enough to prove the local optimality, as we have to show that the stochastic vector $\boldsymbol{Z}^N(t)$ converges to $\boldsymbol{z}^*$ in probability. For that, we introduce the discrete-time version of Kurtz Theorem applied to our problem (see [29]):

Proposition 10. *There exists a neighborhood $\Omega_\delta(\boldsymbol{z}^*)$ of $\boldsymbol{z}^*$ such that if $\boldsymbol{Z}^N(0) = \boldsymbol{z}(0) = \boldsymbol{x} \in \Omega_\delta(\boldsymbol{z}^*)$, then for any $\mu > 0$ and finite time horizon T, there exist positive constants C_1 and C_2 such that*

$$P_{\boldsymbol{x}}\Big(\sup_{0\leq t<T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}(t)|| \geq \mu\Big) \leq C_1 exp(-NC_2) \qquad (47)$$

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

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

Lemma 3. *If $\boldsymbol{Z}^N(0) = \boldsymbol{x} \in \Omega_\delta(\boldsymbol{z}^*)$, 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,*

$$P_{\boldsymbol{x}}\Big(\sup_{T_0\leq t<T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}^*|| \geq \mu\Big) \leq s_1 exp(-Ns_2) \qquad (48)$$

Proof. See appendix N. □

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

Proposition 11. *If the initial state is in the set $\Omega_\delta(\boldsymbol{z}^*)$, then*

$$\lim_{T\to\infty}\lim_{N\to\infty}\frac{C_T^N(\boldsymbol{x})}{N} = \frac{C^{RP,N}}{N} \qquad (49)$$

Proof. See appendix O □

VIII. Global asymptotic optimality

In this section, we prove that from any initial state $\boldsymbol{x}$, 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 $\boldsymbol{Z}^N(t)$. 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. *The state defined as for each class k, $\boldsymbol{z}^k = (1, 0, \cdots, 0)$ and denoted by $\boldsymbol{z}_0$[1], is reachable from any initial state using the Whittle's Index policy.*

Proof. See appendix P □

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

Lemma 5. *By applying the Whittle's Index policy, the steady state $\boldsymbol{z}^*$ is reachable from the state $\boldsymbol{z}_0$.*

[1] $\boldsymbol{z}_0$ can be seen as the system's state where all queues are in the queue state 0

Proof. Considering our system model, 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 $\boldsymbol{z}_0$ to $\boldsymbol{z}^*$ which lasts only one time slot. Consequently, $\boldsymbol{z}^*$ is reachable from $\boldsymbol{z}_0$. □

From this lemma above, the state $\boldsymbol{z}^*$ is reachable from any state, which means that $\boldsymbol{z}^*$ 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 $\boldsymbol{z}^*$ 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 $\Omega_\epsilon(\boldsymbol{z}^*)$. This will ensure that there is a path to enter a neighborhood $\Omega_\epsilon(\boldsymbol{z}^*)$ from any initial state. However, it is important to ensure that the time to enter $\Omega_\epsilon(\boldsymbol{z}^*)$ should not scale up with N. For that, we give the following assumption which will be later justified via numerical studies in Section IX.

Assumption 1. *We assume that the expected time to enter a neighborhood of $\boldsymbol{z}^*$ from any initial state $\boldsymbol{x}$ does not depend on the number of queues N. In other words, for all N the time to enter a neighborhood $\Omega_\epsilon(\boldsymbol{z}^*)$ denoted by $\Gamma^N_{\boldsymbol{x}}(\epsilon)$ is bounded by a constant T_{b_ϵ}.*

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

Lemma 6. *Under Assumption 1, and for any ϵ, we have that:*

$$\lim_{N\to+\infty} P(\boldsymbol{Z}^N(\infty) \in \Omega_\epsilon(\boldsymbol{z}^*)) = 1 \tag{50}$$

Proof. See Lemma 6 in [21]. □

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

$$\lim_{T\to\infty} \frac{C_T^N(\boldsymbol{x})}{N} = \sum_{k=1}^{K}\sum_{i=0}^{L} a_k \mathbb{E}\left[Z_i^{k,N}(\infty)\right] d(i) N \tag{51}$$

Theorem 3. *Under assumption 1, and for any initial state, we have that:*

$$\lim_{N\to+\infty} \lim_{T\to\infty} \frac{C_T^N(\boldsymbol{x})}{N} = \frac{C^{RP,N}}{N} \tag{52}$$

Proof. See appendix Q □

IX. Numerical Results

In this section, we provide numerical results that confirm the asymptotic optimality of the developed Whittle's Index Policy. To that extent, we consider the two following scenarios:

1) 2 classes with the following characteristics:
   - $R_1 = 11$ and $R_2 = 10 \times R_1 = 110$
   - $\alpha = 1/2$
   - $L = 10$
   - $\gamma_1 = \gamma_2 = 1/2$
   - $a_1 = \frac{10*2}{(R_1-1)}$
   - $a_2 = \frac{10*2}{R_2-1}$
   - $C_d = 3$
2) 3 classes with the following characteristics:
   - $R_1 = 11$, $R_2 = 50$ and $R_3 = 110$
   - $\alpha = 1/2$
   - $L = 10$
   - $\gamma_1 = \gamma_2 = \gamma_3 = 1/3$
   - $a_1 = \frac{10*2}{(R_1-1)}$
   - $a_2 = \frac{10*2}{R_2-1}$
   - $a_3 = \frac{10*2}{R_3-1}$
   - $C_d = 3$

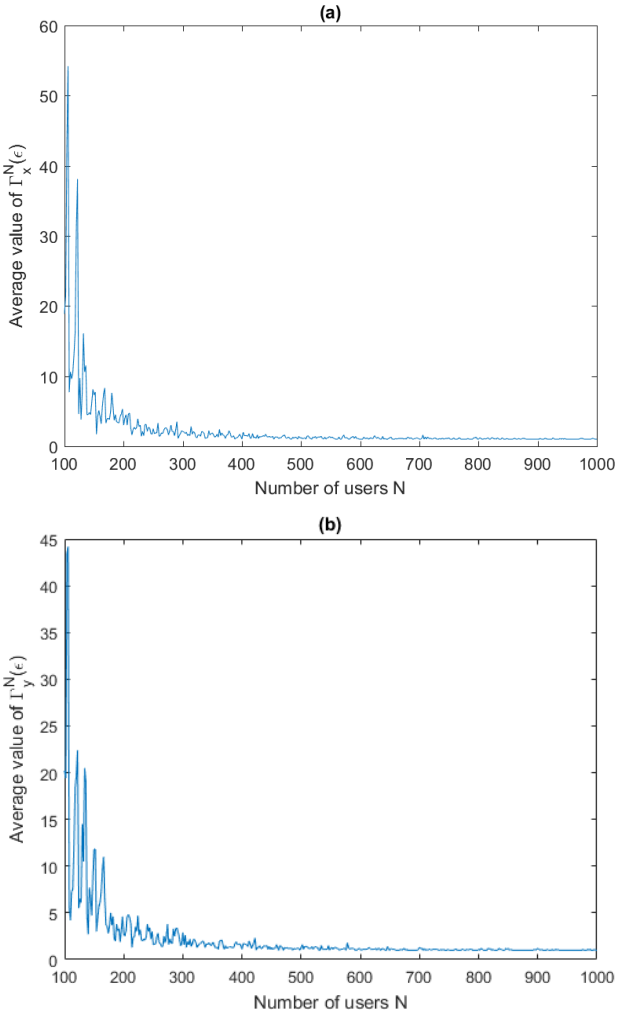

Figure 3: Hitting Time of $\Omega_\epsilon(\boldsymbol{z}^*)$ in function of N: (a) $\boldsymbol{Z}^N(0) = \boldsymbol{x}$, (b) $\boldsymbol{Z}^N(0) = \boldsymbol{y}$

We also consider two initial states x and y such that all the queues are equal to 0 and L respectively.

1) Verification of Assumption 1: We plot in Figure 3, the evolution of the time needed to enter a neighborhood $\Omega_\epsilon(\boldsymbol{z}^*)$ (i.e. hitting time of $\Omega_\epsilon(\boldsymbol{z}^*)$) with respect to N, given that ϵ 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 $\boldsymbol{x}$ and $\boldsymbol{y}$. This implies that the hitting time is bounded for large values of N which consolidates Assumption 1.

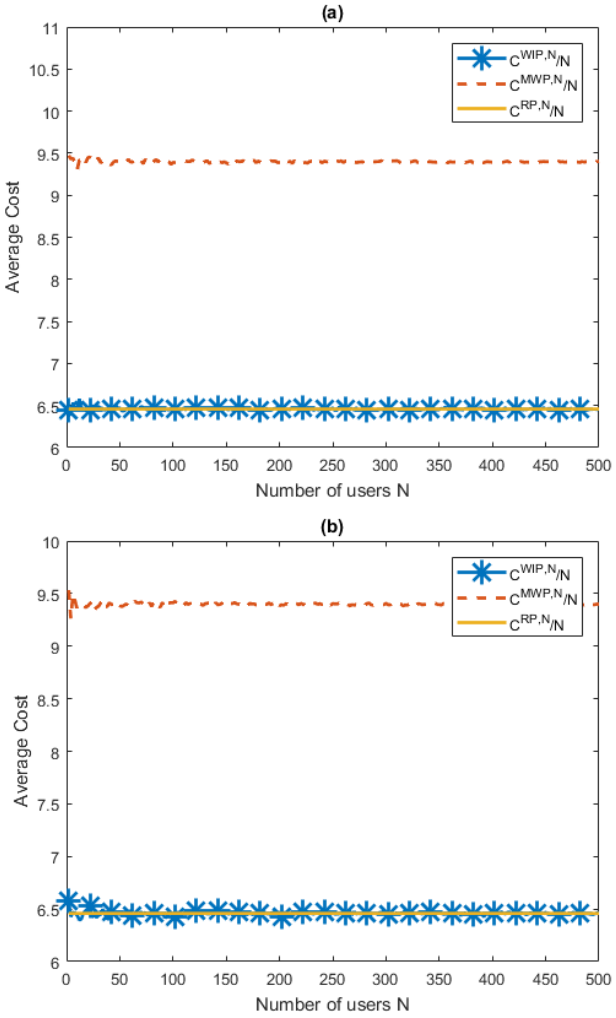

Figure 4: Performance evaluation of Whittle's Index policy for 2 classes

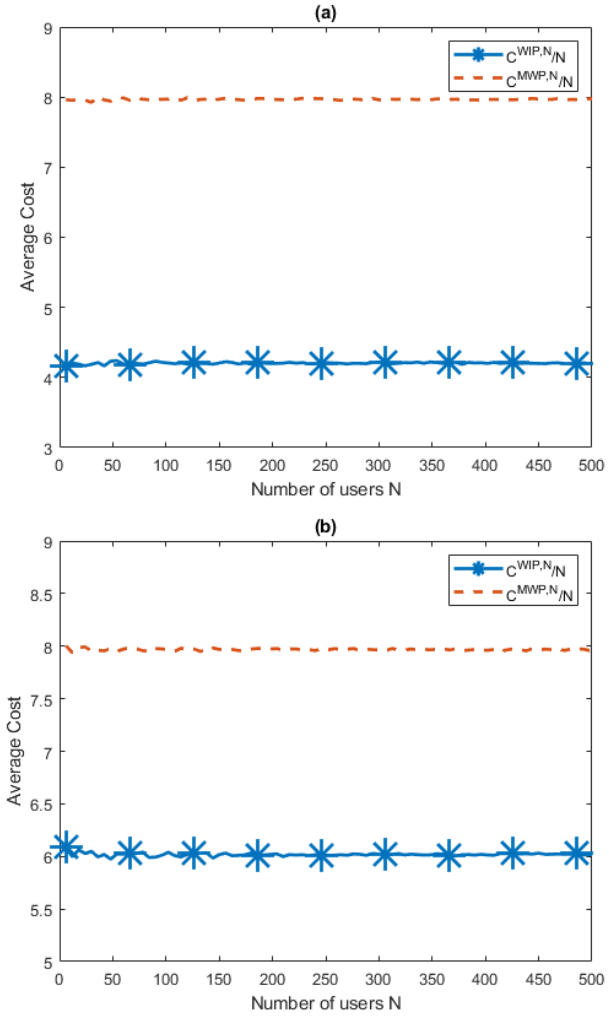

Figure 5: Performance evaluation of Whittle's Index policy for 3 classes

2) 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. $\lim_{T\to\infty} C_T^N(\boldsymbol{x})/N = C^{WIP,N}/N$, 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 $a_k d(q_i^k(t))$). We also compare for the first scenario, 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.a), (4.b), (5.a) and (5.b) where (a) corresponds to the initial state $\boldsymbol{x}$ and (b) corresponds to the initial state $\boldsymbol{y}$ (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 VI). Lastly, we remark that the solution given by MWP is suboptimal and lacks behind our proposed scheduling scheme.

3) Fairness among users: In order to improve the fairness among the users in the network, one can use the developed Whittle's index policy up to some modifications. To that extent, we introduce in this section a new policy Θ 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's index of state $q_i^k(t)$ and $\overline{D_k}(q_i^k(t)) = \frac{\sum_{u=1}^{t} a_k d(q_i^k(u))}{t}$. 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 π, specifically $C_1^{\pi,N} = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{\gamma_1 N} a_1 d(q_i^1(t)) \mid \boldsymbol{x}, \pi\right]$ and $C_2^{\pi,N} = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{i=1}^{\gamma_2 N} a_2 d(q_i^2(t)) \mid \boldsymbol{x}, \pi\right]$. We plot these quantities over N when $\pi = WIP$ and when $\pi = \Theta$ with respect to N in Figure 6 considering the following network settings:

- $R_1 = 11$ and $R_2 = 12$
- $\alpha = 1/2$

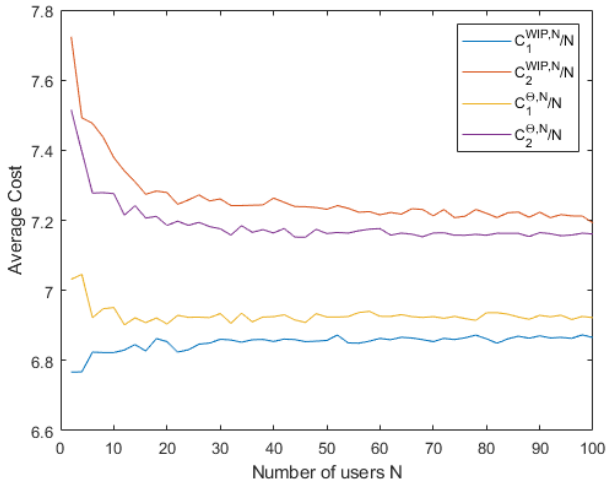

Figure 6: Evaluation of $C_1^{\pi,N}$ and $C_2^{\pi,N}$ in function of N under Policy Θ and Whittle's Index policy WIP

- $L = 10$
- $\gamma_1 = \gamma_2 = 1/2$
- $a_1 = \frac{10*2}{(R_1-1)}$
- $a_2 = \frac{10*2}{R_2-1}$
- $C_d = 3$

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.

X. Conclusion

In this paper, we have studied the problem of users and channels scheduling under dynamic 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. We have formulated a Lagrangian relaxation of the optimization problem and 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 proved its asymptotic local and global optimality when the numbers of users and channels are large enough. 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.

References

[1] S. Kriouile, M. Larranaga, , and M. Assaad, "Whittle index policy for multichannel scheduling in queueing systems," in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 1–6.

[2] M. Deghel, M. Assaad, M. Debbah, and A. Ephremides, "Queueing stability and csi probing of a tdd wireless network with interference alignment," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 547–576, 2018.

[3] A. Destounis, M. Assaad, M. Debbah, and B. Sayadi, "Traffic-aware training and scheduling for the 2-user miso broadcast channel," in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 1376–1380.

[4] ——, "Traffic-aware training and scheduling for miso wireless downlink systems," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2574–2599, 2015.

[5] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE transactions on automatic control*, vol. 37, no. 12, pp. 1936–1948, 1992.

[6] ——, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466–478, 1993.

[7] M. J. Neely, "Optimal energy and delay tradeoffs for multiuser wireless downlinks," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3095–3113, 2007.

[8] L. Georgiadis, M. J. Neely, L. Tassiulas *et al.*, "Resource allocation and cross-layer control in wireless networks," *Foundations and Trends® in Networking*, vol. 1, no. 1, pp. 1–144, 2006.

[9] 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.

[10] 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.

[11] P. Ansell, K. D. Glazebrook, J. Niño-Mora, and M. O'Keeffe, "Whittle's 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.

[12] C. Buyukkoc, P. Variaya, and J. Walrand, "c mu rule revisited." *Adv. Appl. Prob.*, vol. 17, no. 1, pp. 237–238, 1985.

[13] M. Larrañaga, "Dynamic control of stochastic and fluid resource-sharing systems," Ph.D. dissertation, 2015.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] P. Jacko, "Value of information in optimal flow-level scheduling of users with markovian time-varying channels," *Performance Evaluation*, vol. 68, no. 11, pp. 1022–1036, 2011.

[19] F. Cecchi and P. Jacko, "Nearly-optimal scheduling of users with markovian time-varying transmission rates," *Performance Evaluation*, vol. 99, pp. 16–36, 2016.

[20] U. Ayesta, M. Erausquin, and P. Jacko, "A modeling framework for optimizing the flow-level scheduling with time-varying channels," *Performance Evaluation*, vol. 67, no. 11, pp. 1014–1029, 2010.

[21] 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.

[22] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. B. Shroff, "Exploiting channel memory for joint estimation and scheduling in downlink networks," in *INFOCOM, 2011 Proceedings IEEE*. IEEE, 2011, pp. 3056–3064.

[23] M. Larrañaga, M. Assaad, A. Destounis, and G. S. Paschos, "Asymptotically optimal pilot allocation over markovian fading channels," *IEEE Transactions on Information Theory*, 2017.

[24] 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.

[25] S. M. Ross, *Introduction to stochastic dynamic programming*. Academic press, 2014.

[26] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[27] A. Maatouk, S. Kriouile, M. Assad, and A. Ephremides, "On the optimality of the whittle's index policy for minimizing the age of information," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1263–1277, 2020.

[28] Y.-P. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2634–2638.

[29] T. G. Kurtz, "Strong approximation theorems for density dependent markov chains," *Stochastic Processes and their Applications*, vol. 6, no. 3, pp. 223–240, 1978.

[30] J. Gittins, K. Glazebrook, and R. Weber, *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

[31] K. P. Papadaki and W. B. Powell, "Exploiting structure in adaptive dynamic programming algorithms for a stochastic batch service problem," *European Journal of Operational Research*, vol. 142, no. 1, pp. 108–127, 2002.

[32] 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.

[33] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of applied probability*, vol. 25, no. A, pp. 287–298, 1988.

APPENDIX A
PROOF OF PROPOSITION 1

We consider the Bellman Equation (10). By summing the RHS and the LHS of Equation (10), for all k and i we obtain:

$$\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}[V_i^k(q_i^k)+\theta_i^k] = \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\min_{s_i^k}\{C_k(q_i^k,s_i^k) + \sum_{q_i^{'k}} Pr(q_i^{'k}|q_i^k,s_i^k)V_i^k(q_i^{'k})\}$$

$$= \min_{\mathbf{s}}\{\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}[C_k(q_i^k,s_i^k)+\sum_{q_i^{'k}} Pr(q_i^{'k}|q_i^k,s_i^k)V_i^k(q_i^{'k})]\}, \tag{53}$$

where $\mathbf{s} = (s_1^1,\dots,s_{\gamma_1 N}^1,\dots,s_1^K,\dots,s_{\gamma_k N}^K)$. We also have that:

$$\begin{aligned} Pr(\mathbf{q}'|\mathbf{q},\mathbf{s}) &= \sum_{q_i^{'k}} Pr(\mathbf{q}'|\mathbf{q},\mathbf{s},q_i^{'k})Pr(q_i^{'k}|\mathbf{q},\mathbf{s}) \\ &= \sum_{q_i^{'k}} Pr(\mathbf{q}'|\mathbf{q},\mathbf{s},q_i^{'k})Pr(q_i^{'k}|q_k^i,s_k^i), \end{aligned} \tag{54}$$

for all $\mathbf{q} = (q_1^1,\dots,q_{\gamma_1 N}^1,\dots,q_1^K,\dots,q_{\gamma_K N}^K)$ and $\mathbf{q}' = (q_1^{'1},\dots,q_{\gamma_1 N}^{'1},\dots,q_1^{'K},\dots,q_{\gamma_K N}^{'K})$. Since $Pr(q_i^k|\mathbf{q},\mathbf{s})$ only depends on the decision taken with respect to user i in class k, we obtain:

$$\begin{aligned} &\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\sum_{q_i^{'k}} Pr(q_i^{'k}|q_i^k,s_i^k)V_i^k(q_i^{'k}) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\sum_{\mathbf{q}'}\sum_{q_i^{'k}} Pr(\mathbf{q}'|\mathbf{q},\mathbf{s},q_i^{'k})Pr(q_i^{'k}|q_i^k,s_i^k)V_i^k(q_i^{'k}) \\ &= \sum_{\mathbf{q}'} Pr(\mathbf{q}'|\mathbf{q},\mathbf{s})\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} V_i^k(q_i^{'k}) \end{aligned} \tag{55}$$

From the previous equations we obtain:

$$\begin{aligned} &\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} V_i^k(q_i^k) + \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\theta_i^k \\ &= \min_{\mathbf{s}}\,[\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} C_k(q_i^k,s_i^k) + \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\sum_{q_i^{'k}} Pr(q_i^{'k}|q_i^k,s_i^k)V_i^k(q_i^{'k})] \\ &= \min_{\mathbf{s}}\,[\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} C(q_i^k,s_i^k) + \sum_{\mathbf{q}'} Pr(\mathbf{q}'|\mathbf{q},\mathbf{s})\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} V_i^k(q_i^{'k})] \end{aligned} \tag{56}$$

According to [25, Chapter 2, Theorem 2.1], it exists a unique function $\tilde{\boldsymbol{V}}$ and a constant θ that resolve the equation (9). Subsequently, since we have found a bounded function $\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} V_i^k(q_i^k)$, and a constant $\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\theta_i^k$ that satisfy also the equation (9), then $\tilde{\boldsymbol{V}}(\mathbf{q}) = \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} V_i^k(q_i^k)$ and $\theta = \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}\theta_i^k$. 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.

APPENDIX B
PROOF OF PROPOSITION 2

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 π_{i-j}.

2) $j > n$:

The optimal decision in this case is to transmit, 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 π_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 (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. 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 $Pr(L \leq A \leq R-1) = \sum_{k=L}^{R-1} Pr(A = k) = (R - L)\pi_L = (R - L)\rho$.

APPENDIX C
PROOF OF PROPOSITION 3

We have that:

$$u(i) = \sum_{j=0}^{n} p_n(j,i)u(j) + \sum_{j=n+1}^{L} p_n(j,i)u(j) \tag{57}$$

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 2:

$$u(i) = \sum_{j=0}^{n} \pi_{i-j} u(j) + \sum_{j=n+1}^{L} \pi_i u(j) \tag{58}$$

By definition of π given in Definition 2, we have that:

$$u(i) = \sum_{j=0}^{\min(i,n)} \rho u(j) \; + \sum_{j=n+1}^{L} \rho u(j) \tag{59}$$

Now, in order to prove Proposition 3 for this case, we will distinguish between three 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:

$$u(i) = \sum_{0}^{n} \rho u(j) + \sum_{n+1}^{L} \rho u(j) \tag{60}$$

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:

$$\begin{aligned} u(i-1) &= \sum_{0}^{i-1} \rho u(j) + \sum_{n+1}^{L} \rho u(j) \\ u(i-1) &= \sum_{0}^{i} \rho u(j) + \sum_{n+1}^{L} \rho u(j) - \rho u(i) \\ u(i-1) &= u(i) - \rho u(i) \end{aligned} \tag{61}$$

  By induction assumption, we have that $u(i) = \rho(1-\rho)^{n-i}$. To that extent, replacing the expression of $u(i)$ in (61), 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) $n = L$:

$$u(i) = \sum_{j=0}^{L} p_L(j,i) u(j) \tag{62}$$

For $i \leq L-1$:
According to Proposition 2, we have:

$$u(i) = \sum_{j=0}^{L} \pi_{i-j} u(j) \tag{63}$$

By definition of π, we get:

$$u(i) = \sum_{0}^{i} \rho u(j) \tag{64}$$

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:

$$\begin{aligned} u(i+1) &= \sum_{0}^{i+1} \rho u(j) \\ &= \sum_{0}^{i} \rho u(j) + \rho u(i+1) \\ &= 0 + \rho u(i+1) \\ u(i+1) &= 0 \end{aligned} \tag{65}$$

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.

Appendix D
Proof of Proposition 4

As mentioned previously in Remark 2, we denote $\sum_{q=0}^{L} a u^n(q) d(q)$ by a_n and $\sum_{q=0}^{n} u^n(q)$ by b_n. Before proving the proposition, we give two useful lemmas.

Lemma 7. *Considering a_{j-1}, a_j, a_{j+1} and b_{j-1}, b_j, b_{j+1}, such that $b_{j-1} < b_j < b_{j+1}$.*

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:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \tag{66}$$

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:*

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \tag{67}$$

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:

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \tag{68}$$

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:*

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \tag{69}$$

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:*

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \leq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \leq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \tag{70}$$

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:*

$$\frac{a_j - a_{j-1}}{b_j - b_{j-1}} \geq \frac{a_{j+1} - a_{j-1}}{b_{j+1} - b_{j-1}} \geq \frac{a_{j+1} - a_j}{b_{j+1} - b_j} \tag{71}$$

Proof. See appendix E ■

Lemma 8. *The largest minimizer at step j in Algorithm 1 satisfies $n_j = \min\{k : b_k = b_{n_j}\}$*

Proof. See appendix F. ■

- Indexability:
  We consider n_1 and n_2 the optimal thresholds of the problem (10) 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 that 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 (8) when $W = W_1$ and $W = W_2$ respectively, then:

  $$a_{n_1} - W_1 b_{n_1} \leq a_{n_2} - W_1 b_{n_2} \tag{72}$$

  $$a_{n_1} - W_2 b_{n_1} \geq a_{n_2} - W_2 b_{n_2} \tag{73}$$

  This implies:

  $$W_2(b_{n_1} - b_{n_2}) \leq a_{n_1} - a_{n_2} \leq W_1(b_{n_1} - b_{n_2}) \tag{74}$$

  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 8, b_k is necessarily strictly less than b_{n_0}. Thus, by using Lemma 7 (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 8).
    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 7 (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 7 (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 7 (first case), $\frac{a_{n_{j+1}} - a_k}{b_{n_{j+1}} - b_k} \leq W_{j+1}$. Therefore, for all $k \leq n_{j+1}$, $a_k - b_k W_{j+1} \geq a_{n_{j+1}} - b_{n_{j+1}} W_j$.

  As a 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 E
PROOF OF LEMMA 7

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:

$$\frac{a_{j+1}-a_{j-1}}{b_{j+1}-b_{j-1}} = \frac{a_{j+1}-a_j}{b_{j+1}-b_{j-1}} + \frac{a_j-a_{j-1}}{b_{j+1}-b_{j-1}} \tag{75}$$

$$\geq \frac{(a_j-a_{j-1})(b_{j+1}-b_j)}{(b_j-b_{j-1})(b_{j+1}-b_{j-1})} + \frac{a_j-a_{j-1}}{b_{j+1}-b_{j-1}} \tag{76}$$

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

$$\frac{a_{j+1}-a_{j-1}}{b_{j+1}-b_{j-1}} \geq \frac{a_j-a_{j-1}}{b_j-b_{j-1}}[\frac{b_{j+1}-b_j+b_j-b_{j-1}}{b_{j+1}-b_{j-1}}] \tag{77}$$

$$= \frac{a_j-a_{j-1}}{b_j-b_{j-1}} \tag{78}$$

For the RHS inequality:

$$\frac{a_{j+1}-a_{j-1}}{b_{j+1}-b_{j-1}} = \frac{a_{j+1}-a_j}{b_{j+1}-b_{j-1}} + \frac{a_j-a_{j-1}}{b_{j+1}-b_{j-1}} \tag{79}$$

$$\leq \frac{a_{j+1}-a_j}{b_{j+1}-b_{j-1}} + \frac{(a_{j+1}-a_j)(b_j-b_{j-1})}{(b_{j+1}-b_j)(b_{j+1}-b_{j-1})} \tag{80}$$

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

$$\frac{a_{j+1}-a_{j-1}}{b_{j+1}-b_{j-1}} \leq \frac{a_{j+1}-a_j}{b_{j+1}-b_j}[\frac{b_{j+1}-b_j+b_j-b_{j-1}}{b_{j+1}-b_{j-1}}] \tag{81}$$

$$= \frac{a_{j+1}-a_j}{b_{j+1}-b_j} \tag{82}$$

APPENDIX F
PROOF OF LEMMA 8

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:

$$\frac{a_{n_j}-a_{n_{j-1}}}{b_{n_j}-b_{n_{j-1}}} \leq \frac{a_i-a_{n_{j-1}}}{b_i-b_{n_{j-1}}} \tag{83}$$

$$\frac{a_{n_j}-a_{n_{j-1}}}{b_{n_j}-b_{n_{j-1}}} \leq \frac{a_i-a_{n_{j-1}}}{b_{n_j}-b_{n_{j-1}}} \tag{84}$$

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.

APPENDIX G
PROOF OF PROPOSITION 5

When $n \in [0, L-1]$, we have $a_n - a_{n-1} = a\rho C_d + a[\rho[(L-n) - (C_d + L + R)(1-\rho)^n] + 1]$. To that extent, we denote by $g(n)$ the function: $a\rho C_d + a[\rho[(L-n) - (C_d + L + R)(1-\rho)^n] + 1]$ and we show that $g(.)$ is positive for $n \in [0, L]$. To that end, we give the second derivative of $g(.)$:

$$g''(n) = a[-\rho(ln(1-\rho))^2(C_d + L + R)(1-\rho)^n] \tag{85}$$

It is clear from the above equation that $g''(.)$ is non positive. Hence, $g(.)$ is concave function with n. That is, for all $n \in [0, L]$, $g(n) \geq \min\{g(0), g(L)\}$. Thereby, our task will be to demonstrate that $g(0)$ and $g(L)$ are both positive. In fact, $g(0) = 0 \geq 0$. While for $n = L$, it requires more technical analysis to establish the desired result. To that end, we decompose the function $g(.)$ into two functions $t(\cdot)$ and $f(.)$ such that:

$$g(n) = t(n) + f(n)$$

where

$$t(n) = a\rho C_d - a\rho(1-\rho)^n C_d$$

and

$$f(n) = a[\rho[(L-n) - (L+R)(1-\rho)^n] + 1]$$

We show that $t(L)$ and $f(L)$ are both positive.
We have $t(L) = a\rho C_d(1-(1-\rho)^L) \geq 0$.
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,:

$$v(n) = a[\rho[-1 + \rho(L+R)(1-\rho)^n]] \tag{86}$$

knowing that $v(n)$ is lower bounded by $a[\rho[-1+\rho(L+R)(1-\rho)^L]]$ for $n \in [0, L-1]$, then:

$$f(L) \geq \sum_{n=0}^{L-1} a[\rho[-1 + \rho(L+R)(1-\rho)^L]]$$
$$= a[L\rho[-1 + \rho(L+R)(1-\rho)^L]] \tag{87}$$

Therefore:

$$f(L) = a[1 - \rho(L+R)(1-\rho)^L]$$
$$\geq a[-L\rho[1 - \rho(L+R)(1-\rho)^L]] \tag{88}$$

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. As a consequence, since $g(L)$ is the sum of two positive terms, then $g(L)$ is also positive. Providing that $g(n) \geq \min\{g(0), g(L)\}$ for $n \in [0, L]$, then $g(n) \geq 0$. Hence for $n \in [0, L-1]$,

$$a_n \geq a_{n-1} \tag{89}$$

We still have to show that $a_L - a_{L-1} \geq 0$. In fact:

$$a_L - a_{L-1} = aC_d[1-(1-\rho)^L] + a[R - (L+R)(1-\rho)^L]$$
$$= aC_d[1-(1-\rho)^L] + a[R(1 - \rho(L+R)(1-\rho)^L)]$$
$$\geq 0 \tag{90}$$

Thus, combining the two results (89) and (90), we end up with the desired result.

APPENDIX H
PROOF OF LEMMA 1

At $W = x_{i,j}$, $y^i(W) = y^j(W)$, i.e.:

$$\sum_{q=0}^{L} au^i(q)d(q) - W\sum_{q=0}^{i} u^i(q) = \sum_{q=0}^{L} au^j(q)d(q) - W\sum_{q=0}^{j} u^j(q)$$

$$\sum_{q=0}^{L} au^i(q)d(q) \ - \sum_{q=0}^{L} au^i(q)d(q) = W\sum_{q=0}^{i} u^i(q) - \ W\sum_{q=0}^{j} u^j(q) \tag{91}$$

$$\sum_{q=0}^{L} au^i(q)d(q) \ - \sum_{q=0}^{L} au^i(q)d(q) = W[\sum_{q=0}^{i} u^i(q) - \sum_{q=0}^{j} u^j(q)] \tag{92}$$

Hence

$$W = \frac{\sum_{q=0}^{L} au^i(q)d(q) - \sum_{q=0}^{L} au^j(q)d(q)}{\sum_{q=0}^{i} u^i(q) - \sum_{q=0}^{j} u^j(q)} \tag{93}$$

APPENDIX I
PROOF OF THEOREM 2

In order to prove this theorem, we introduce the following useful lemmas.

Lemma 9. *$x_{n,n-1}$ is strictly increasing with n*

Proof. We have for all $n \in [0, L-1]$:

$$x_{n+1,n} - x_{n,n-1} = W(n+1) - W(n) = \frac{a\rho(L + C_d - n)}{(1-\rho)^{n+1}} > 0$$

That concludes the proof. ■

Lemma 10. *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 $k < s \leq L$:

$$\frac{a_s - a_{k-1}}{b_s - b_{k-1}} > \frac{a_k - a_{k-1}}{b_k - b_{k-1}} \tag{94}$$

Proof. We fix certain $k \in [0, L-1]$, we prove the result by induction:
for $s = k+1$

$$\begin{aligned}\frac{a_{k+1} - a_{k-1}}{b_{k+1} - b_{k-1}} =& \frac{a_{k+1} - a_{k-1} - a_k + a_k}{b_{k+1} - b_{k-1}} \\ =& \frac{a_{k+1} - a_k}{b_{k+1} - b_{k-1}} + \frac{a_k - a_{k-1}}{b_{k+1} - b_{k-1}} \\ >& \frac{(a_k - a_{k-1})(b_{k+1} - b_k)}{(b_k - b_{k-1})(b_{k+1} - b_{k-1})} \\ &+ \frac{(a_k - a_{k-1})(b_k - b_{k-1})}{(b_k - b_{k-1})(b_{k+1} - b_{k-1})}\end{aligned} \tag{95}$$

where the strict inequality comes from the lemma's assumptions. Therefore, we have that:

$$\begin{aligned}\frac{a_{k+1} - a_{k-1}}{b_{k+1} - b_{k-1}} &> \frac{a_k - a_{k-1}}{b_k - b_{k-1}}[\frac{b_{k+1} - b_k}{b_{k+1} - b_{k-1}} + \frac{b_k - b_{k-1}}{b_{k+1} - b_{k-1}}] \\ &= \frac{a_k - a_{k-1}}{b_k - b_{k-1}}\end{aligned} \tag{96}$$

By induction, we consider that the inequality (94) is true for certain s strictly higher than k. The inequality below is then verified for $s+1$:

$$\begin{aligned}\frac{a_{s+1} - a_{k-1}}{b_{s+1} - b_{k-1}} =& \frac{a_{s+1} - a_{k-1} - a_s + a_s}{b_{s+1} - b_{k-1}} \\ =& \frac{a_{s+1} - a_s}{b_{s+1} - b_{k-1}} + \frac{a_s - a_{k-1}}{b_{s+1} - b_{k-1}} \\ >& \frac{(a_k - a_{k-1})(b_{s+1} - b_s)}{(b_k - b_{k-1})(b_{s+1} - b_{k-1})} \\ &+ \frac{(a_k - a_{k-1})(b_s - b_{k-1})}{(b_k - b_{k-1})(b_{s+1} - b_{k-1})} \\ =& \frac{a_k - a_{k-1}}{b_k - b_{k-1}}[\frac{b_{s+1} - b_s}{b_{s+1} - b_{k-1}} + \frac{b_s - b_{k-1}}{b_{s+1} - b_{k-1}}] \\ =& \frac{a_k - a_{k-1}}{b_k - b_{k-1}}.\end{aligned} \tag{97}$$

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's indices, we denote by j the step j described in the algorithm.
According to the same algorithm, to show that $x_{j,j-1}$ is the Whittle's index at 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 9, $W(j) < W(j+1) < .... < W(L)$. 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 10, 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 a consequence, the Whittle's index of state j according to Algorithm 1 is effectively $W(j) = \frac{a_j - a_{j-1}}{b_j - b_{j-1}} = x_{j,j-1}$.

APPENDIX J
PROOF OF PROPOSITION 6

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's index less than W. Then, according to the definition of Whittle's 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 2, $W_j^k = x_{j,j-1}^k$ which is the point for which if $W = x_{j,j-1}^k$, we have $\sum_{q=0}^{L} a_k u_k^j(q)d(q) - W\sum_{q=0}^{j} u_k^j(q) = \sum_{q=0}^{L} a_k u_k^{j-1}(q)d(q) - W\sum_{q=0}^{j-1} u_k^{j-1}(q)$. That means, according to Equation (22), 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's 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 $W_0^k < \cdots < W_{j-1}^k < W_j^k = W$, 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.

Appendix K
Proof of Proposition 7

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:

$$\max_W \min_{\phi\in\Phi} f(W,\phi) \leq \min_{\phi\in\Phi} \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} a_k d(q_i^k(t)) \mid \boldsymbol{q}(0),\phi\right] \tag{98}$$

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:

$$\begin{aligned}\max_W \min_\phi & f(W,\phi) \\ =&\max_W\{\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}[\min_{l_k\in[0,L]}\{\sum_{q=0}^{L} a_k u_k^{l_k}(q)d(q) - W\sum_{q=0}^{l_k} u_k^{l_k}(q)\}] \\ &+ W(1-\alpha)N\}\end{aligned} \tag{99}$$

with ϕ being the threshold policy that corresponds to $l(W)$ computed using Proposition 6 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: $\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}[-W\sum_{q=0}^{l_k(W^*)} u_k^{l_k(W^*)}(q)] + W(1-\alpha)N = \sum_{k=1}^{K}\gamma_k N[-W(1-\sum_{i=l_{k+1}(W^*)}^{L} u_k^{l_k(W^*)}(i))] + W(1-\alpha)N = -NW + \sum_{k=1}^{K}\gamma_k NW\sum_{i=l_{k+1}(W^*)}^{L} u_k^{l_k(W^*)}(i) + W(1-\alpha)N = -NW + \alpha N + WN - \alpha N = 0$. Hence, we get:

$$\begin{aligned}\min_\phi & f(W^*,\phi) \\ &= f(W^*, l(W^*)) = \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}[\sum_{q=0}^{L} a_k u_k^{l_k(W^*)}(q)d(q)] \\ &= \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} a_k d(q_i^k(t)) \mid \boldsymbol{q}(0), l(W^*)\right]\end{aligned} \tag{100}$$

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). Moreover, according to the inequality (98), we have that for all policy ϕ that satisfies the constraint and belong to Φ:

$$\begin{aligned} f(W^*,&l(W^*)) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N}[\sum_{q=0}^{L} a_k u_k^{l_k(W^*)}(q)d(q)] \\ &= \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} a_k d(q_i^k(t)) \mid \boldsymbol{q}(0), l(W^*)\right] \\ &= \min_\phi f(W^*,\phi) \\ &\leq \max_W \min_\phi f(W,\phi) \\ &\leq \min_\phi \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=1}^{\gamma_k N} a_k d(q_i^k(t)) \mid \boldsymbol{q}(0), \phi\right].\end{aligned} \tag{101}$$

We deduce that the solution of the relaxed problem is of type threshold-based policy $l(W^*)$ with W^* satisfies $\alpha = \sum_{k=1}^{K}\gamma_k\sum_{i=l_{k+1}(W^*)}^{L} u_k^{l_k(W^*)}(i)$.

Appendix L
Proof of Proposition 8

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 6, 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 with $l(W)$, and take discrete values from 1 to 0. According to Proposition 6, 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 6, 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 θ 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 α. When $\theta=1$, the threshold policy is $l_m(W_p^m)$ and the total average time in active mode is less than α.
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 θ, then there exists θ^* 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 θ^*, the constraint (4) is satisfied with equality:

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

Appendix M
Proof of Proposition 9

We derive the eigenvalues of $\boldsymbol{Q}$.
The matrix $\boldsymbol{Q}$ is of the form:

$$\begin{bmatrix} Q_1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & Q_2 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & & \\ A_1 & A_2 & \cdots & Q_m & \cdots & A_{K-1} & A_K \\ \vdots & & & \ddots & & \vdots & \\ 0 & 0 & \cdots & \cdots & \cdots & Q_{K-1} & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & Q_K \end{bmatrix} \tag{102}$$

The characteristic polynomial of $\boldsymbol{Q}$ is the product of the characteristic polynomial of each matrix Q_k:

$$\chi_{\boldsymbol{Q}}(\lambda) = \prod_{k=1}^{K}\chi_{Q_k}(\lambda) \tag{103}$$

Therefore, the set of $\boldsymbol{Q}$'s eigenvalues denoted by $\mathrm{Sp}(\boldsymbol{Q})$ is composed by the eigenvalues of the matrices Q_k. Specifically: $\mathrm{Sp}(\boldsymbol{Q}) = \bigcup_k \mathrm{Sp}(\boldsymbol{Q})$. To that extent, it is sufficient to find the eigenvalues of each matrix Q_k to deduce those of $\boldsymbol{Q}$. To that end, we distinguish between two cases:

1) $k \neq m$: The characteristic polynomial of the matrix Q_k is defined as follows:

$$\chi_{Q_k} = \det(Q_k - \lambda I) \tag{104}$$

where $I \in \mathbb{R}^{(L+1)\times(L+1)}$ is the identity matrix. 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 and c_i the row i and column i respectively of the determinant. We also denote by $a_{i,j}$ the element in row i and column j of the matrix Q_k. For $i=0$ till l_k-1, we add to the row r_L the sum of the rows r_i for $0 \leq i \leq l_k-1$. In other words:

$$r_L = r_L + \sum_{i=0}^{l_k-1} r_i \tag{105}$$

After doing so, we execute the following operation in order to have zeros for the elements $a_{L,0}$ to a_{L,l_k-1}:

$$c_i = c_i - c_L \quad i = 0, \ldots, l_k - 1 \tag{106}$$

As a result, $\chi_{Q_k}(\lambda)$ will be the determinant of the matrix G_k reported in Table II. Since G_k is an upper triangular matrix, the determinant will be simply the product of diagonal elements of matrix G_k. Hence, the determinant of G_k will be equal to $(-\lambda)^{l_k}$ times the $(-\lambda)^{L-l_k+1}$.

As a consequence, the determinant is equal to:

$$\chi_{Q_k}(\lambda) = (-\lambda)^L \tag{107}$$

2) $k = m$: The characteristic polynomial of the matrix Q_m is defined as follows:

$$\chi_{Q_k} = \det(Q_m - \lambda I) \tag{108}$$

The matrix $Q_m - \lambda I$ is a lower triangular matrix. Therefore, its determinant will be simply equal to:

$$\chi_{Q_m}(\lambda) = (-\lambda)^{L-l_m}(\rho_m - \lambda)^{l_m} \tag{109}$$

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 ρ_m which are strictly less than 1. Consequently, in both cases, whether $k \neq m$ or $k = m$, the norms of all eigenvalues of Q_k are strictly less than 1. Hence, for $\lambda \in \mathrm{Sp}(\boldsymbol{Q}) \Rightarrow |\lambda| < 1$.

$$
Q_k = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ \vdots \\ l_k-2 \\ l_k-1 \\ l_k+1 \\ \vdots \\ L-1 \\ L \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} 0 & 1 & \cdots & l_k-2 & l_k-1 & l_k+1 & l_k+2 & \cdots & L-1 & L \end{array} \\
\begin{pmatrix}
\rho_k & 0 & \cdots & 0 & 0 & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
\rho_k & \rho_k & \ddots & 0 & 0 & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \rho_k & \rho_k & \vdots & \vdots \\
\vdots & & \ddots & \rho_k & 0 & \vdots & & & \vdots & \vdots \\
\rho_k & \cdots & \cdots & \rho_k & \rho_k & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
\vdots & 0 & 0 & \vdots & \vdots & \vdots & 0 & 0 & \vdots & \vdots \\
0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
-l_k\rho_k & (1-l_k)\rho_k & \cdots & -2\rho_k & -\rho_k & -l_k\rho_k & \cdots & \cdots & -l_k\rho_k & -l_k\rho_k
\end{pmatrix}
\end{array}
$$

$$
Q_m = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ \vdots \\ l_m-2 \\ l_m-1 \\ l_m+1 \\ \vdots \\ L-1 \\ L \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} 0 & 1 & \cdots & l_m-2 & l_m-1 & l_m+1 & l_m+2 & \cdots & L-1 & L \end{array} \\
\begin{pmatrix}
\rho_m & 0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
\rho_m & \rho_m & \ddots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & 0 & 0 & \vdots & \vdots \\
\vdots & & \ddots & \rho_m & 0 & \vdots & & & \vdots & \vdots \\
\rho_m & \cdots & \cdots & \rho_m & \rho_m & 0 & \cdots & \cdots & 0 & 0 \\
0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
\vdots & 0 & 0 & \vdots & \vdots & \vdots & 0 & 0 & \vdots & \vdots \\
0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\
-l_m\rho_m & (1-l_m)\rho_m & \cdots & -2\rho_m & -\rho_m & 0 & \cdots & \cdots & 0 & 0
\end{pmatrix}
\end{array}
$$

Table I: The expressions of the matrices Q_k for $k \neq m$ and Q_m

$$
G_k = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ \vdots \\ l_k-2 \\ l_k-1 \\ l_k+1 \\ \vdots \\ \vdots \\ L-1 \\ L \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} 0 & 1 & \cdots & l_k-2 & l_k-1 & l_k+1 & l_k+2 & \cdots & L-1 & L \end{array} \\
\begin{pmatrix}
-\lambda & -\rho_k & \cdots & -\rho_k & -\rho_k & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
0 & -\lambda & \ddots & -\rho_k & -\rho_k & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
\vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \rho_k & \rho_k & \vdots & \vdots \\
\vdots & & \ddots & -\lambda & -\rho_k & \vdots & & & \vdots & \vdots \\
0 & \cdots & \cdots & 0 & -\lambda & \rho_k & \cdots & \cdots & \rho_k & \rho_k \\
0 & \cdots & \cdots & 0 & 0 & -\lambda & 0 & \cdots & 0 & 0 \\
\vdots & 0 & 0 & \vdots & \vdots & \vdots & -\lambda & 0 & \vdots & \vdots \\
\vdots & 0 & 0 & \vdots & \vdots & \vdots & 0 & \ddots & 0 & \vdots \\
0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & 0 & -\lambda & 0 \\
0 & 0 & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & -\lambda
\end{pmatrix}
\end{array}
$$

Table II: The expressions of the matrix G_k for $k \neq m$

APPENDIX N
PROOF OF LEMMA 3

We take $0 < \epsilon < \mu$, $\boldsymbol{z}(t)$ converges to $\boldsymbol{z}^*$, i.e. there exists T_0 such that for all $t \geq T_0$, $||\boldsymbol{z}(t) - \boldsymbol{z}^*|| \leq \epsilon$. Hence:

$$
\begin{aligned}
&P_x(\sup_{T_0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}^*|| \geq \mu) \\
&\leq P_x(\sup_{T_0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}(t)|| + ||\boldsymbol{z}(t) - \boldsymbol{z}^*|| \geq \mu) \\
&\leq P_x(\sup_{T_0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}(t)|| \geq \mu - \epsilon) \\
&\leq P_x(\sup_{0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}(t)|| \geq \mu - \epsilon) \qquad (110)
\end{aligned}
$$

Using Proposition 10, there exists s_1 and s_2 such that:

$$
P_x(\sup_{0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}(t)|| \geq \mu - \epsilon) \leq s_1 exp(-N s_2). \quad (111)
$$

Therefore:

$$
P_x(\sup_{T_0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}^*|| \geq \mu) \leq s_1 exp(-N s_2). \quad (112)
$$

APPENDIX O
PROOF OF PROPOSITION 11

We recall that $\boldsymbol{Z}^N(t)$ represents the proportion vector at time t under Whittle's Index policy.
Replacing $C^{RP,N}$ by its expression given in Section VI 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 $\boldsymbol{z}^*$), then the difference between $C_T^N(\boldsymbol{x})$ and $C^{RP,N}$ can be expressed as:

$$C_T^N(\boldsymbol{x}) - C^{RP,N} = \Big| \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} a_k Z_i^{k,N}(t) i N \mid \boldsymbol{x}\right] - \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} a_k z_i^{k,*} i N\right] \Big| \quad (113)$$

We divide all by N

$$\begin{aligned}
&\frac{C_T^N(\boldsymbol{x})}{N} - \frac{C^{RP,N}}{N} \\
&= \Big| \frac{1}{T}\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} \mathbb{E}(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i) \Big| \\
&\leq \Big| \frac{1}{T}\sum_{t=0}^{T_0-1}\sum_{k=1}^{K}\sum_{i=0}^{L} \mathbb{E}(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i) \Big| \\
&+ \Big| \frac{1}{T}\sum_{t=T_0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} \mathbb{E}(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i) \Big| \\
&\leq \frac{T_0(L+C_d)(L+1)}{T}\sum_{k=1}^{K} a_k \gamma_k \\
&+ \Big| \frac{1}{T}\sum_{t=T_0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} \mathbb{E}(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i) \Big|
\end{aligned} \quad (114)$$

We have that the function $f: \boldsymbol{z} \to \sum_{k=1}^{K}\sum_{i=0}^{L} a_k z_i^k d(i)$ is lipchitz and continuous, then for an arbitrary small ϵ, there exists μ such that if $\big|\big|\boldsymbol{z} - \boldsymbol{z}^*\big|\big| < \mu$, then $|f(\boldsymbol{z}) - f(\boldsymbol{z}^*)| < \epsilon$.

We denote Y_N the event $\sup_{T_0 \leq t < T} ||\boldsymbol{Z}^N(t) - \boldsymbol{z}^*|| \geq \mu$, we proceed to bound the second term:

$$\begin{aligned}
&\Big| \frac{1}{T}\sum_{t=T_0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} E(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i) \Big| \\
&\leq P_x(Y_N)\frac{1}{T}\sum_{t=T_0}^{T-1}\mathbb{E}\left[\Big|\sum_{k=1}^{K}\sum_{i=0}^{L}(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i)\Big| \Big| Y_N\Big|\right] \\
&+ (1 - P_x(Y_N))\frac{1}{T}\mathbb{E}\left[\Big|\sum_{t=T_0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L}(a_k Z_i^{k,N}(t) d(i)) - a_k z_i^{k,*} d(i)\Big| \Big| \overline{Y_N}\right] \\
&\leq \frac{(T-T_0)(L+C_d)(L+1)}{T}\sum_{k=1}^{K} a_k \gamma_k P_x(Y_N) + (1 - P_x(Y_N))\epsilon.
\end{aligned} \quad (115)$$

where the above inequality comes from the fact that $|a_k Z_i^{k,N}(t) d(i) - a_k z_i^{k,*} d(i)| \leq 2\gamma_k a_k d(i)$. According to Lemma 3, we have $\lim_{N\to\infty} P_x(Y_N) = 0$, then:

$$\begin{aligned}
&\lim_{N\to\infty} \Big| \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} a_k Z_i^{k,N}(t) d(i) N \mid \boldsymbol{x}\right] \\
&\quad - \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} a_k z_i^{k,*} d(i) N\right] \Big| \\
&\leq \frac{T_0(L+C_d)(L+1)}{T}\sum_{k=1}^{K} a_k \gamma_k + \epsilon
\end{aligned} \quad (116)$$

This inequality is true $\forall \epsilon > 0$, then:

$$\begin{aligned}
&\lim_{N\to\infty} \Big| \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} a_k Z_i^{k,N}(t) d(i) N \mid \boldsymbol{x}\right] \\
&\quad - \frac{1}{T}\mathbb{E}\left[\sum_{t=0}^{T-1}\sum_{k=1}^{K}\sum_{i=0}^{L} a_k z_i^{k,*} d(i) N\right] \Big| \\
&\leq \frac{T_0(L+C_d)(L+1)}{T}\sum_{k=1}^{K} a_k \gamma_k
\end{aligned} \quad (117)$$

Finally we have:

$$\lim_{T\to\infty}\lim_{N\to\infty} \frac{C_T^N(\boldsymbol{x})}{N} - \frac{C^{RP,N}}{N} = 0 \quad (118)$$

APPENDIX P
PROOF OF LEMMA 4

We consider any initial state $(\boldsymbol{z}^1, \boldsymbol{z}^2, \cdots, \boldsymbol{z}^K)$, 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 $\boldsymbol{z}_0$. For that purpose, we divide the queues into $\frac{1}{\alpha}$ groups denoted by $G_1, \cdots, G_\alpha$ such that G_k contains a proportion α of queues with the highest Whittle's indices among all queues of the system excluding those of the groups $G_1, G_2, \cdots, G_{k-1}$ 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 2, the Whittle's index of state 0 is equal to 0 whatever the value of the class. While according to the same Proposition, the Whittle's index of state n strictly higher than 0, is strictly greater than 0 for any class k. Therefore, regardless of the class, the Whittle's 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 $\boldsymbol{z}_0$. That concludes the proof.

APPENDIX Q
PROOF OF THEOREM 3

$$\lim_{T\to\infty}\frac{C_T^N(\boldsymbol{x})}{N}-\frac{C^{RP,N}}{N}$$
$$=\sum_{k=1}^{K}\sum_{i=0}^{L}a_k\mathbb{E}\left[Z_i^{k,N}(\infty)\right]d(i)-\sum_{k=1}^{K}\sum_{i=0}^{L}a_k z_i^{k,*}d(i) \tag{119}$$

We have the function $f:\ \boldsymbol{z}\to\sum_{k=1}^{K}\sum_{i=0}^{L}a_k z_i^k d(i)$ is lipchitz and continuous, then for an arbitrary small ϵ, there exists μ such that if $||\boldsymbol{z}-\boldsymbol{z}^*||<\mu$, then $|f(\boldsymbol{z})-f(\boldsymbol{z}^*)|<\epsilon$.

We denote U_N the event $\sup||\boldsymbol{Z}^N(\infty)-\boldsymbol{z}^*||\geq\mu$, then :

$$\Big|\sum_{k=1}^{K}\sum_{i=0}^{L}a_k\mathbb{E}\left[Z_i^{k,N}(\infty)\right]d(i)-\sum_{k=1}^{K}\sum_{i=0}^{L}a_k z_i^{k,*}d(i)\Big|$$
$$\leq P(U_N)\mathbb{E}\left[\Big|\sum_{k=1}^{K}\sum_{i=0}^{L}(a_k Z_i^{k,N}(\infty)d(i))-a_k z_i^{k,*}d(i)\Big|\Big|U_N\right]$$
$$+(1-P(U_N))\mathbb{E}\left[\Big|\sum_{k=1}^{K}\sum_{i=0}^{L}(a_k Z_i^{k,N}(\infty)d(i))-a_k z_i^{k,*}d(i)\Big|\Big|\overline{U_N}\right]$$
$$\leq(L+C_d)(L+1)\sum_{k=1}^{K}a_k\gamma_k P(U_N)+(1-P(U_N))\epsilon \tag{120}$$

According to Lemma 6, we have $\lim_{N\to\infty}P(U_N)=0$, then:

$$\lim_{N\to\infty}\Big|\sum_{k=1}^{K}\sum_{i=0}^{L}a_k\mathbb{E}\left[Z_i^{k,N}(\infty)\right]d(i)-\sum_{k=1}^{K}\sum_{i=0}^{L}a_k z_i^{k,*}d(i)\Big|\leq\epsilon \tag{121}$$

This is true for any ϵ. Finally we have:

$$\lim_{N\to\infty}\Big|\lim_{T\to\infty}\frac{C_T^N(\boldsymbol{x})}{N}-\frac{C^{RP,N}}{N}\Big|=0 \tag{122}$$

That completes the proof.