Scheduling users in drive-thru Internet: a multi-armed bandit approach

Thi Thuy Nga Nguyen, Urtzi Ayesta, Balakrishna Prabhu

Scheduling users in drive-thru Internet: a multi-armed bandit approach*

Thi Thuy Nga NGUYEN[1,3], Urtzi AYESTA[2], and Balakrishna PRABHU[3]

[1]Continental Digital Service in France, Toulouse, France
[2]CNRS-IRIT, Univ. Toulouse, Toulouse, France
[3]LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

April 18, 2019

Abstract

We consider the problem of allocating a wireless channel to mobile users moving on a straight road. The objective is to maximize a given function of the total data transmitted. We develop a model within the multi-armed bandit framework and we formulate an optimization problem under the constraint that only one user can be served at a time. We solve the relaxed optimization problem, in which one user is served on the average, for which the solution is given by Whittle's index policy, and we derive a heuristic policy for the original optimization problem using Whittle's index policy as well. We evaluate numerically to see how well the heuristic algorithm performs in various settings, including the dynamic scenario with arrivals of new users, and in the presence of heterogeneous users.

Index terms— Markov Decision Process, restless multi-armed bandit problem, Whittle's index, scheduling, drive-thru internet

1 Introduction

Drive-thru internet has seen a recent resurgence due to an increase in demand for high-speed internet access from mobile users [1, 2, 3]. One of the early works in this area [4], introduced this concept as WLAN for mobile users. Until then, local networks such as IEEE 802.11 b/a/g were primarily targeted towards stationary users. Using measurements, they concluded that WLAN hotspots could be a viable technology for high-speed internet access in different mobility settings. The capability of present generation hand-held mobile devices to support high bandwidth applications such as video transfers has reignited interest in this technology.

A typical scenario for drive-thru internet studies is a WiFi hotspot or access point (AP) that servers users moving along a straight line as shown in Figure 1. For example, these users can be cars or pedestrians moving on (or along) a long avenue. Various question related to link-layer scheduling and resource allocation [5, 6], MAC layer retransmissions [7], message scheduling using network coding [8] have recently been investigated by taking into account the specific mobility pattern of the drive-thru internet systems.

In this paper, we revisit the multi-class scheduling problem for Markovian queues [9, 10] in the context of a drive-thru internet. Consider users of different classes (i.e, different mean service requirements) moving along a straight line in the coverage area of an AP (Fig. 1). Users enter the coverage range from the left and leave from the right. In each time-slot, the AP has

*This is an extended version of the work that appears in Wiopt 2019.

to determine which user to serve in order to maximize a given long-term objective. The AP can serve at most one user in each time-slot. Users receive a rate depending upon their distance from the AP: users who are closer have higher rate (as shown in Fig. 1). The trade-off is between serving users with a higher rate and users who leave first.

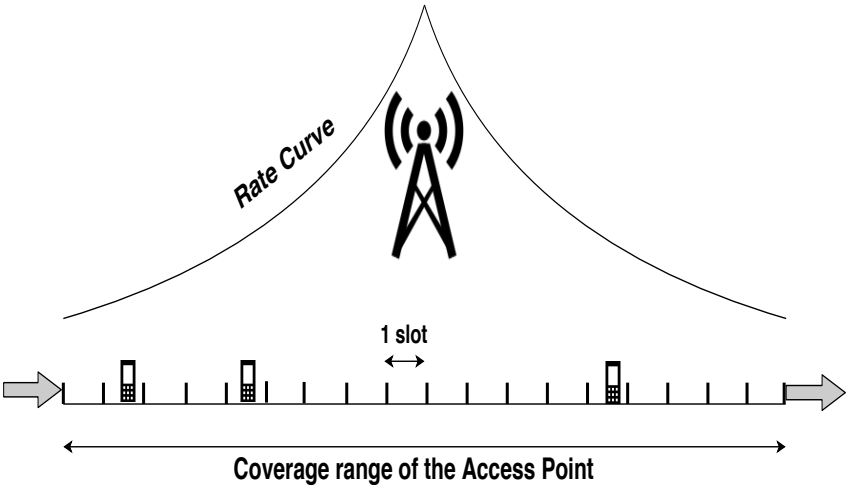

Figure 1: A drive-thru internet network.

1.1 Contributions

The problem as a Markov Decision Process [11] whose solution can be computed numerically for small number of users but becomes computationally intractable for large instances. We shall rely on the multi-armed bandit approach of Whittle [12, 13] to obtain a heuristic based on the Whittle index. In general, the Whittle indices are not easy to compute ([14, 15]) but given the special mobility structure (users move one spatial-slot to the right in each time-slot), we are able to obtain a simple computational procedure for these indices. The indices are obtained as a function of the position and the class of the user. The heuristic then serves the user with the highest index. It will be shown that this index is not the same as the greedy algorithm (or the $c\mu$-rule [16]) that assigns the channel to the user with the largest product of channel rate and mean service requirement. Several numerical experiments will be presented to show the performance improvements of the proposed heuristic with respect to the greedy policy. In particular, the improvements are more pronounced when there are more classes of users.

1.2 Related work

In [8] the authors develop an information-theoretic formula for the total amount of information that a vehicle can receive (for only one user) when it passes the broadcast zone of a BS. Vehicles moving in a road may have blind zones in which they might not receive signal from the base station. To circumvent this problem, they propose and analyse the benefits of cooperative scheme for joint V2V and V2I communications in order to improve the system capacity.

The Gittins' index is optimal for, what is known in the bandit literature as, the 'rested' bandit problem, i.e, in which the bandits do not change state if they are not served. Moving users, on the other hand, are 'restless', that is, users change state (or position) even when they are not served. For the bandits that restless, Whittle developed a relaxation based method which we shall employ to attack the problem. This method is presented in [12]-[10].

As explained in the introduction, one of the main contributions of this paper is the closed-form calculation of Whittle's index, which enables to develop a simple heuristic, known as

Whittle's index policy (WIP), for the original problem. WIP is a particular instance of so-called *index policies*, that is, the solution to the stochastic control problem is characterized by an *index*, a function that depends on the state of a single job, that determines which action to take.

Index policies have long be known in scheduling theory. In general, the solution to a scheduling problem will be a complex function of all the input parameters and the number of competing jobs. In practice such problems can be solved only for very specific instances. Remarkably, in some cases, a so-called *index policy* is optimal. Two well-known examples of this situation are the so-called $c\mu$-rule (optimal in a single server with linear holding costs and exponential service times [16]) and Shortest-Remaining-Processing-Time (SRPT, optimal in a single server where the remaining service times are known [17].)

The above two optimality results ($c\mu$ and SRPT) can be cast in the framework of Multi-Armed Bandit Problems (MABP), a broad class of resource allocation problems for which Index policies are known to be optimal. A MABP is a particular case of a Markov Decision Process: at every decision epoch the scheduler needs to select one *bandit*, and an associated reward is accrued. The state of this selected bandit evolves stochastically, while the state of all other bandits remains *frozen*. The scheduler knows the state of all bandits and aims at maximizing the total average reward. In a ground-breaking result Gittins showed that the optimal policy that solves a MABP is an index rule, nowadays commonly referred to as Gittins' index policy [18]. Thus, for each bandit, one calculates Gittins' index, which depends only on its own current state and stochastic evolution. The optimal policy activates in each decision epoch the bandit with highest current index.

Despite its generality, in multiple cases of practical interest the problem cannot be cast as a MABP. For example, mobility of users directly invalidates the requirement that non-selected bandits remain *frozen*. In a seminal work [19], Whittle introduced the so-called Restless Bandit Problem (RBP), a generalization of the standard MABP in which all bandits might evolve over time according to a stochastic kernel that depends on whether the bandit is made active. RBP provides a powerful modeling framework, but its solution has in general a complex structure that might depend on the entire state-space description. In fact, it is known that RBP are PSPACE-hard even in its deterministic variant [20], and is typically attributed to suffer from the curse of dimensionality.

Whittle considered a relaxed version of the problem (where the restriction on the number of *active* bandits needs to be respected on average only, and not in every decision epoch), and showed that the solution to the relaxed problem is of index type, referred to as *Whittle's index*. Whittle then defined a heuristic for the original problem, referred to as Whittle's index policy, where in every decision epoch the bandit with highest Whittle index is selected. It has been shown that the Whittle index policy performs strikingly well, see [21] for a discussion, and is asymptotically optimal under certain conditions, see [9, 22].

In addition to resource allocation problems, Whittle's index has been applied in a wide variety of cases, including recommendation systems, website morphing and pharmaceutical trials, [18, Chapter 6]. In the last years many researchers have applied Whittle's index approach for the opportunistic scheduling in wireless networks, see [15, 23] and references therein.

In order to calculate Whittle's index there are two main difficulties: first, one needs to establish a technical property known as *indexability*, and second, the calculation of the Whittle index itself might be involved or even infeasible. Indexability has been established for a wide variety of cases, but in many cases, see for example [14], the calculation of the index is done numerically.

1.3 Organization

The rest of the paper is organized as follows. Section 2 formally describes the general setting and casts the problem as an MDP. It also proposes the simpler model of no arrivals. Section 3

states the main result on the indexability of the simpler model of no arrivals and gives a simple numerical procedure for the computation of the indices. The heuristic Whittle-index policy based upon the main result in presented in Section 4. This section also contains numerical comparison of the proposed policy with other policies. The conclusions and further research directions appear in Section 5. Most of the proofs have been moved to the Appendix for improved readability.

2 Problem formulation

Consider an AP with a coverage range of length L (see Fig. 1). The users enter the coverage range from left, move at a constant velocity, and leave from the right. Every Δ time units the AP has to decide which user to serve. Let v be the velocity of the users. Then, the coverage range can be divided into spatial-slots on length $v\Delta = \sigma$. Let $\mathcal{S} = \{1, 2, \ldots, N\}$, where $N = L/\sigma$, denote the set of spatial-slots with the convention that slot 1 is the leftmost slot. The length of the time-slot is assumed to be much smaller than the coverage range of the AP (in the order of hundreds of meters). The scheduling decisions are made every 10-20 ms during which a car inside a city would move a distance of less than a metre.

In each time-slot, the AP has to choose *at most* one user (or a spatial-slot) to serve, that is, its set of actions is $\mathcal{A} = \{e_i\}_{i\in\mathcal{S}}$ with e_i being the unit vector for the ith coordinate. The user that will be chosen by the AP in a time-slot will depend upon the data rate of the users currently in range as well as the class of these users. The data rate received by a user in spatial-slot s depends on the distance between the AP and s. Users that are closer to the AP will get a higher rate than the users that are closer to the end points. We shall assume that the Signal-to-Noise Ratio (SNR) has a polynomial decay:

$$SNR(s) = \frac{C_1}{d(s)^\gamma},$$

and that the data rate in slot s, $C(s)$ can be obtained using the the Shannon law:

$$C(s) = C_2 \log(1 + SNR(s)). \tag{1}$$

For more information on these formulae, we refer to [8]. The amount of data that is transmitted in a time-slot, $r(s)$, to a user served at rate $C(s)$ will thus be $C(s)\Delta$.

Assumption 1. *The function $r(s)$ is unimodal with maximum at $s = N/2$ (assuming N is even). It is non-decreasing on the left and non-increasing on the right.*

The assumption is quite natural and is satisfied by the rate function derived from (1).

The total volume of data requested by user i is assumed to have volume $D_{i,b}$ to transfer. Here b is the class of user i and $b \in \mathcal{B} := \{1, 2, ...B\}$, where B is the number of classes. We shall assume that, for each b, $D_{i,b}$ are independently and exponentially distributed with rate η_b. The probability that a user of class-b who is served in slot s finishes its data transfer in that slot is $1 - \exp(-\eta_b r(s))$. The assumption of exponential data volumes ensures that this probability is independent of past allocations.

In each time-slot, users arrive in spatial-slot 1 according to a categorical distribution on $\mathcal{B} \cup \{0\}$. The outcome 0 corresponds to no arrival in that time-slot. The probability that a user of class-b arrives in time-slot will be denoted p_b for $b \in \mathcal{B}$. If $\sum_b p_b < 1$, then there is non-zero probability of there being no new arrival in a time-slot.

2.1 Objective

For a given policy π of the AP, let $S^\pi(t) \in (\{0, 1\} \times B)^{\mathcal{S}}$ be the stochastic process that tells whether a spatial-slot is occupied by a user or not and tells the class of the user if it is occupied.

The objective of the AP is:

$$\max_{\pi} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} R^{\pi}(S(t), a(S(t), t)), \tag{OBJEC}$$

where $a(S(t), t) \in \mathcal{A}$ is the action prescribed by policy π in the state S in time-slot t. And $R^{\pi}(s, a)$ is reward one-step after chosing action a for state s which is described in this below assumption.

Assumption 2. *The reward in a time-slot is sum of the rewards of each user, where the reward of a user is a strictly positive and increasing function of the rate if it is served and* 0 *if it is not served.*

From now to present easier , reward one-step of a user is its departure probability in that slot, which is a strictly positive and increasing function of that user's rate.

From the assumptions on the data volumes and the arrival process, it can be seen that this problem is a classical average-cost MDP [11]. There exists an optimal stationary (time-independent) policy that can be computed numerically. The drawback of this formulation, however, is that the number of states in any practical scenario is too large to allow numerical computation. As mentioned in the Introduction, for time-slots of 10–20 ms and a coverage length of 100–200 m, the number of spatial-slots, N, of the order of a thousand. The state space of $S(t)$ will have $\approx B2^{|N|} = B2^{1000}$ elements making the problem intractable. Even for 20–30 spatial-slots, the problem is not computationally tractable in reasonable time with the current technology.

Instead of the treating the problem in its full generality, we shall as first step focus on a simplified instance of the problem in which there are no arrivals, that is $p_b = 0$, $\forall b \in \mathcal{B}$. This will allow us to obtain certain heuristics that can be then used for the general problem.

2.2 Finite horizon MDP for problem with no-arrivals

Let there be K users at time 0, and let $X_k(t) \in \mathcal{N} := \mathcal{S} \cup N + 1$ be the position of user-k in time-slot t. The special state $N + 1$ indicates that the user has departed the system either because it has moved out of the coverage range or because its demand has been satisfied. We shall assume that the parameter of the exponential distribution for user-k is η_k. That is, each user could potentially be of a different class.

Since there are no arrivals, the process $S(t)$ can be replaced by the process $\mathbf{X}(t) := (X_1(t), \ldots X_K(t))$. Let $a_k(t) \in \{0, 1\}$ denote whether user-k was served in slot t or not, and let $\mathbf{a}(t) := (a_1(t), \ldots, a_K(t))$.

With these definitions, it can be seen that the problem (OBJEC) is equivalent to the following problem finite-horizon MDP when there are no arrivals:

$$\begin{cases} \max_{\pi} & \frac{1}{N+1} \sum_{t=0}^{N} \sum_{k=1}^{K} \mathbb{E}_x^{\pi}(R_k(X_k(t), a_k(t)) \\ \text{subject to} & \\ & \sum_{k=1}^{K} a_k(t) \leq 1,\ t = 0, 1, 2, ..., N, \\ & a_k(t) = \{0, 1\},\ \forall k, t \end{cases} \tag{NOARR}$$

Here $\mathbf{X}(0) = x$ is the initial position of the users and $R_k(x, a)$ is the reward obtained (i.e., data transferred) by the user-k when action $a_k(t)$ is taken. The constraints on the actions indicate that at most one user can be served in a time-slot. Further, the horizon of the problem can be constrained to N since all users would have left the coverage range by that time.

In general, finite-horizon problem need not have optimal policies that are stationary. However, problem (NOARR) is a particular case known as the stochastic shortest path problem ([24], e.g.) for which under certain assumptions there exists a stationary optimal policy which is the solution of Bellman's equation.

Lemma 1. *Problem* (NOARR) *admits a stationary optimal policy that satisfies Bellman's equation.*

Proof. We need to check that Assumptions 1 and 2 in [24] are satisfied by (NOARR). Assumption 1 requires the existence of a stationary *proper* policy, that is, a stationary policy that takes the process to its terminal state. Since all users will leave the coverage range in $N+1$ time-slots irrespective of the actions of the AP, all policies (stationary or not) are *proper* and there are no *improper* policies. Thus, this requirement is met. It also requires the terminal state (state $N+1$ in our problem) to have 0 cost which again is satisfied. Finally, any improper policy should have an initial state from which its cost is infinite. Since there are no improper policies, all the requirements of Assumption 1 are met. Assumption 2 is satisfied by finite-state and finite-action problems which is true for (NOARR). From Proposition 2 in [24], we can conclude that there exists a stationary optimal policy that satisfies Bellman's equation. □

This result will be important later on when we shall derive an heuristic based on Whittle's index.

With some abuse of notation, let $r_x = r(x)$. For a user-k, given $a_k(t) = a$, $X_k(t)$ has the transition probabilities:

$$\mathbf{P}_k(y|x,a) = \begin{cases} ae^{-r_x\eta_k} + (1-a), & y = x+1, x \neq N+1; \\ a(1-e^{-r_x\eta_k}), & y = N+1, x \neq N+1; \\ 1, & y = N+1, x = N+1; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

It follows from the above that:

- The dynamics of each user is Markovian and is independent of that of the other users conditioned on the action. Further, each user can change state whether it is served or not.
- The reward function is decomposable into sum of rewards of the individual users.

Problem (NOARR) is thus an instance of the Restless Multi-armed Bandit (MAB)framework considered by Whittle [12, 13]. The bandits in that framework correspond to users in our problem. Since the users can change state even when they are not served, the users are restless bandits[1]. Problem (NOARR) is a finite-horizon problem whereas most of the work is in the literature on MABs is on the infinite-horizon setting. The fact that (NOARR) is also a classical stochastic shortest path problem allows us to use results of the infinite-horizon setting for the present problem as well.

Whittle proposed a relaxation which allows to decompose (NOARR) into K sub-problem thus reducing the dimension of the problem considerably (from N^K to N). Each of these K sub-problems can sometimes be solved analytically, and from these solutions one can obtain an heuristic called Whittle's index policy, that is quite easy to implement. It has been observed on several problems that a heuristic based on Whittle's index performance very well in practice [25, 26, 27] and is in fact asymptotically optimal when the number of bandits becomes large [22].

We shall follow the approach of Whittle to obtain an heuristic policy for (NOARR).

3 Whittle's relaxation and indexability

One of the difficulty in solving (NOARR) comes from the constraints that need to be satisfied in each time-slot. To overcome this, Whittle proposed to relax the constraint that exactly at

[1]From now on, we shall use bandits and users interchangeably to mean the same thing. Similarly activating a bandit will mean serving a user.

most one user is served (or active) per time and to replace it by the constraint that at most one user is active in average per time. He then considered the Lagrange relaxation of the problem with relaxed constraints and arrived at K sub-problems—one for each of the K users.

For (NOARR), this approach leads to the problem:

$$\max_{\pi} \frac{1}{N+1} \sum_{t=0}^{N} \sum_{k=1}^{K} \mathbb{E}_x^{\pi}(R_k(X_k(t), a_k(t)) \\ -\nu \frac{1}{N+1} \sum_{t=0}^{N} \sum_{k=1}^{K} \mathbb{E}_x^{\pi}(a_k(t)),$$

where ν is the Lagrange multiplier of the constraint.

Once there are no constraints on the actions, there is no longer any dependence between the users. Following Whittle approach, we use this property to obtain the following K sub-problems:

$$\max_{\pi_k} \sum_{t=1}^{N} \mathbb{E}_{x_k}^{\pi_k}(R_k(X_k(t), a_k(t)) - \nu \sum_{t=1}^{N} \mathbb{E}_{x_k}^{\pi_k}(a_k(t)). \tag{SUBP-k}$$

where π_k is policy of single user-k. Sub-problem-k, (SUBP-k), is associated to user-k when this user is alone in the system and there is penalty ν on the actions. Each (SUBP-k) is again a stochastic shortest path problem which can be solved independently of the other users, and to which the reasoning of Lemma 1 can be applied to argue the existence of an optimal stationary policy.

Intuitively, ν can be seen as the penalty for being active (or being served) because it reduces the reward for taking $a_k > 0$. If $\nu = -\infty$, then it is optimal to activate all the bandits while if $\nu = +\infty$, then the optimal policy is to inactivate the bandits.

From now we concentrate on (SUBP-k), and omit index k in the variables to simplify the notation. For a given ν, any stationary policy, π, can be characterized by its active set $\Omega^{\pi}(\nu) = \{x : a(x) = 1\}$ which is the set of states in which the bandit is active. Let $\Omega^*(\nu) \subset \mathcal{N}$ to be the active set for the optimal policy of (SUBP-k). It can be seen that $\Omega^*(0) = \mathcal{N}$ is the set of all states. This is because there is no penalty for taking $a = 1$ and in each state this action gives at least as much immediate reward as $a = 0$. Similarly, $\Omega^*(\infty) = \emptyset$ since $a = 1$ has too high a penalty.

Definition 1. *For $\nu \in [0, \infty)$, a bandit is said to be indexable if $\Omega^*(\nu)$ is monotonically decreasing in ν, that is $\nu_1 \leq \nu_2 \Leftrightarrow \Omega^*(\nu_1) \supseteq \Omega^*(\nu_2)$.*

The Fig. 3 illustrates indexablity on an example with numerically with $|\mathcal{N}| = 200$ and $\eta = 1/3$, since once a state is in passive zone it never comes back active zone.

If a bandit is indexable, we can define the Whittle index of a state (for more details see [12], [13]).

Definition 2. *Given an indexable bandit, the Whittle index ν_x of a state x, is the largest value of ν such that action active x is optimal in that state. That is, $v_x = \sup\{\nu | x \in \Omega^*(\nu)\}$.*

$$\nu_x \geq \nu \Leftrightarrow x \in \Omega^*(\nu). \tag{3}$$

The index ν_x gives us an indication to how profitable it is to activate the bandit (or serve the user) in state x. If $\nu_x > \nu_y$, it means that it is even with a lower subsidy ν it is profitable to be active in state x than in state y.

This motivates the following *heuristic policy*: given the state, the data rate, and the class of each user in the coverage, the AP serves the user with the highest current Whittle index. For this heuristic to be work, the bandits need to be indexable. In the next section, we show that this is true for the bandits defined by (SUBP-k) and give a relatively cheap method for the computation of the Whittle indices.

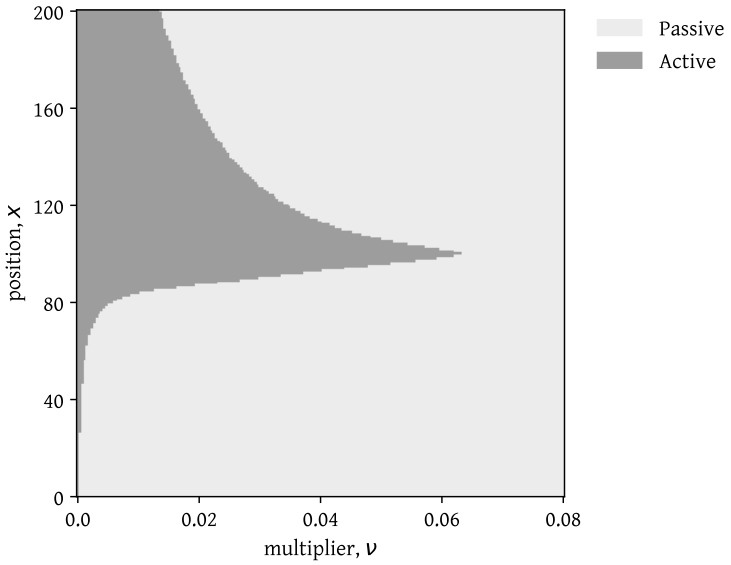

Figure 2: Illustration of indexability.The darker (resp. lighter) region indicates the bandit is active (resp. passive). $\eta = 1/3$ and $N = 200$.

3.1 Indexability

In the rest of the paper, we shall make the following assumptions which will simplify the presentation. The results carry over under the more general conditions mentioned in Assumptions 1 and 2.

Assumption 3. *1. The probability of leaving due to service in a time-slot in state x is approximated by $\eta_b r_x$.*

2. The reward function is $R_k(x, a) = r_x \cdot \eta_k \cdot a$. Here, the factor η_k can be seen as the weight of user-k.

The approximation of probability is justified when the duration of time-slot is small compared to the average time required for service completion. Hence, the probability of departure due to service completion, $1 - e^{-\eta_b r_x}$ can be approximated by $\eta_b r_x$.

For indexability, we shall restrict the domain of ν to $[0, +\infty)$. Denote by $V(x, \nu)$ the value function, i.e.

$$V(x, \nu) = \max_{\pi} \sum_{t=0}^{N} \mathbb{E}_x^{\pi}(R(X(t), a(t)) - \nu \sum_{t=0}^{N} \mathbb{E}_x^{\pi}(a(t)).$$

As mentioned earlier, even though (SUBP-k) is a finite-horizon problem, its value function satisfies Bellman's equation because (SUBP-k) is a classical stochastic shortest path problem (see Lemma 1). Therefore, its value function is the solution of

$$V(x, \nu) = \max_{a \in \{0,1\}} \left((R(x, a) - \nu)a + \sum_{y \in \mathcal{N}} \mathbf{P}(y|x, a) V(y, \nu) \right).$$

Replacing $R(x, a)$ and $\mathbf{P}(y|x, a)$ with values from Assumption 3, the above equation simplifies to:

$$V(x, \nu) = \max_{a \in \{0,1\}} \{(r_x \eta - \nu)a + (1 - r_x \eta a) V(x + 1, \nu)\}. \tag{4}$$

Recall that state $N+1$ is the terminal state in which the user has left, so $V(N+1,\nu)=0$ for any ν. For $x=0,1,...N$, define:

$$\begin{aligned} V^0(x,\nu) &= V(x+1,\nu), \\ V^1(x,\nu) &= (r_x\eta-\nu)+(1-r_x\eta)V(x+1,\nu). \end{aligned}$$

Bellman's equation then becomes $V(x,\nu)=\max_{\{0,1\}}\{V^1(x,\nu),V^0(x,\nu)\}$.

Remark 1. *: $V^1(x,\nu),V^0(x,\nu),V(x,\nu)$ are continuous and non-increasing in ν, since they derived from the maximum of finite number of continuous and non-increasing functions of ν.*

From the definition of indexability (see Definition 2), Whittle's index of state x, is the value of ν such that

$$V^1(x,\nu_x)=V^0(x,\nu_x) \text{ with } \nu\in[0,+\infty), \tag{5}$$

which is equivalent to

$$r_x\eta-\nu=r_x\eta V(x+1,\nu) \text{ with } \nu\in[0,+\infty). \tag{6}$$

We shall prove that for any x, (5) has *exactly one* solution, called ν_x, and thus it is Whittle's index of the state x. The existence and uniqueness of the solution implies indexability. Indeed, if (5) has a unique solution, then due to continuity of $V^1(x,\nu)$ and $V^0(x,\nu)$ in ν it implies that the sign of $V^1(x,\nu)-V^0(x,\nu)$ changes only once in $[0,\infty)$ and this change happens at ν_x. Since $V^1(x,\infty)<V^0(x,\infty)$, we have $V^1(x,\nu)<V^0(x,\nu)$ for $\nu\in[\nu_x,\infty)$ and $V^1(x,\nu)\geq V^0(x,\nu)$ otherwise. This argument will be made formal in Theorem 1 below.

Assume N is even (the arguments of the proof also work when N is odd.) It will be convenient to divide the state-space, $\mathcal{N}$, into two subsets: one on the left of the AP, $\mathcal{N}^-=\{0,1,2,...,N/2-1\}$ and one to the right of the AP (including in front of the AP), $\mathcal{N}^+=\{N/2,N/2+1,...,N+1\}$. For convenience, define $f(x,\Delta)$ for $x\in\mathcal{N}^-,\Delta\in\mathcal{N}^+$ as follows:

$$f(x,\Delta):=\frac{r_x\eta\big(1-\sum_{i=x+1}^{\Delta}r_i\eta\prod_{j=x+1}^{i-1}(1-r_j\eta)\big)}{1-r_x\eta(\sum_{i=x+1}^{\Delta}\prod_{j=x+1}^{i-1}(1-r_j\eta))}, \tag{7}$$

This following theorem shows the indexability and gives the formula for the unique solution and characterizes the behavior of the indices.

Theorem 1 (Indexability)**.** *For each state x, the equation (5) has the unique solution denoted by ν_x. It implies indexability of the bandit k.*
More precise, the index is given in this following formula:

1.1 On the right $x\geq N/2,\nu_x=r_x\eta$, and $\nu_{N/2}>\nu_{N/2+1}>\cdots>\nu_N$.

1.2 On the left $x\leq N/2-1,\nu_x=f(x,\Delta(x))$ where $\Delta(x)\in\mathcal{N}^+$ such that $f(x,\Delta(x))\in[r_{\Delta(x)+1}\eta,r_{\Delta(x)}\eta)$, and $\nu_1<\nu_2<\cdots<\nu_{N/2-1}<\nu_{N/2}$.

We note that the index follows exactly the same pattern as the data rate curve r_x. That is, ν_x is increasing for $x\in\mathcal{N}^-$ and decreasing for $x\in\mathcal{N}^+$. Further, for $x\in\mathcal{N}^+$, it is equal to the probability of departure $r_x\eta$ whereas on the left (i.e., $x\in\mathcal{N}^-$), $\nu_x<r_x\eta$.

In this following proposition, we prove that Whittle's index policy always gives more priority for the state on the right hand side.

Proposition 1. *(Right priority) Suppose r_x is a symmetric about $x=N/2$. If x and y are symmetric ($x+y=N$), x is on the left ($x<N/2$), y is on the right ($y\geq N/2$) then $\nu_x<\nu_y$.*

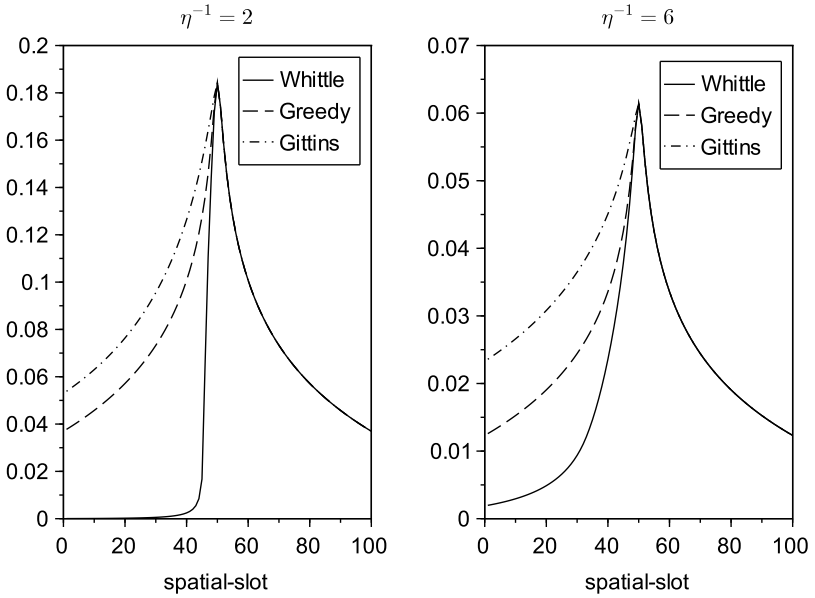

Figure 3: Comparison of the one-step reward curve (Greedy), Gittins' index and the Whittle's index for two different values of η. The three indices are coincide on the right but not coincide on the left.

So if we have two cars that have the same rates but are on the opposite sides of the AP, then Whittle index policy chooses the car on the right hand side.

Fig. 3 illustrates the Whittle index, and compares it with the one-step reward (which is also the departure probability), and Gittin's index curve (which is recalled in the full version). The one-step reward can be seen as the index of the greedy algorithm which chooses the user with the highest one-step reward. The Whittle index gives priority to the users on the right-hand side because they leave the system earlier than users on the left-hand side. Further, the users on the left-hand side will pass through much more favorable channel conditions later on. Thus, one can wait to serve them later and hope to get a better reward.

Whittle's index for more genaral case

So far we consider the case when there one type of cars who leave at the same state (every car moving until state N and then leave the system). In this section we consider the more general case, when there are not only one type of cars, i.e there are some cars who leave before the others because of the structure of the road (for example there is another road across to the considered road). We denote type 1 containing the cars who leave exactly after N, and type 2 containing the cars who leave sooner (they leave after state $N_1 < N$). Denote ν_x^1 and ν_x^2 the Whittle's index of state x in type 1 and type 2 respectively. In this case, Whittle's index changes in this way:

Proposition 2. *$\nu_x^1 \leq \nu_x^2$ for every state $x = 0, 1, ..., N$.*

Intuitively, the cars leaving sooner should have more priority.

4 Whittle-index based policy

Coming back to the original optimization problem, by using the formula in Theorem 1 we can calculate the indices of all vehicles currently present in the coverage range of the AP. Our

proposed policy based on Whittle's method is to allocate the channel to the user that has the highest current Whittle index. Let $\mathcal{K}(t)$ be the set of users present in the coverage range at time-step t. The algorithm for this policy takes as input the current position, the data rate and the mean size of the demand as input. It them computes the Whittle index of each user and selects the one with the highest one. If there are two or more users with the same index, one is chosen arbitrarily.

Algorithm WIP: Heuristic policy based on Whittle indices

for *every time step* t **do**
  Input : Vectors $\mathbf{X}(t)$, $\mathbf{r}_{\mathbf{X}(t)}$, and η
  Output: $\mathbf{a}^*$
  $\mathbf{a}^* \leftarrow 0$
  $i = \arg\max_{k \in \mathcal{K}(t)} \nu_{k,X_k(t)}$
  `/* choose arbitrarily one maximizer, if there is more than one */`
  $a_i^* \leftarrow 1$
end

The proposed policy shall be compared with the following policies.

- *Optimal*: obtained by solving (OBJEC) (or (SUBP-k) depending upon the scenario). The optimal policy can only be computed for small number of time-slots so will not be shown when this number is large.
- *Greedy*: chooses the user with the best one-step reward.
- *Gittin's index*: serves the user with the best Gittin's index.
- *RMS*: gives priority to the right-most user. This mimics the Whittle's index by serving users on the right-hand side. However, it goes a step further and gives priority to users who are leaving first.
- *LMS*: gives priority to the left-most user.

4.1 No arrivals

First, we compare the average reward obtained when there are no arrivals to the system and one class of users. The number of time-slots is $N = 100$, the mean service requirement is $\eta^{-1} = 1$, and the rate curve, r_x, is given in Fig. 3. There are K cars in the system and their initial position is chosen randomly. The total reward for a run is computed and then the experiment is repeated by taking a different initial condition. For various values of K, Table 1 gives the values of the average total reward obtained after averaging over 1000 experiments for different policies. The optimal policy is not evaluated because of the large size of the state-space. WIP

Table 1: Compare algorithms in case of no arrival.

Number of cars	**Whittle**	**Greedy**	**Gittin**	**RMS**	**LMS**
$K = 10$	0.082	0.070	0.069	0.077	0.061
$K = 20$	0.129	0.114	0.113	0.080	0.086
$K = 40$	0.195	0.190	0.189	0.080	0.109
$K = 60$	0.226	0.224	0.224	0.080	0.122

outperforms all the other policies, except for RMS for low values of K. At the two extreme

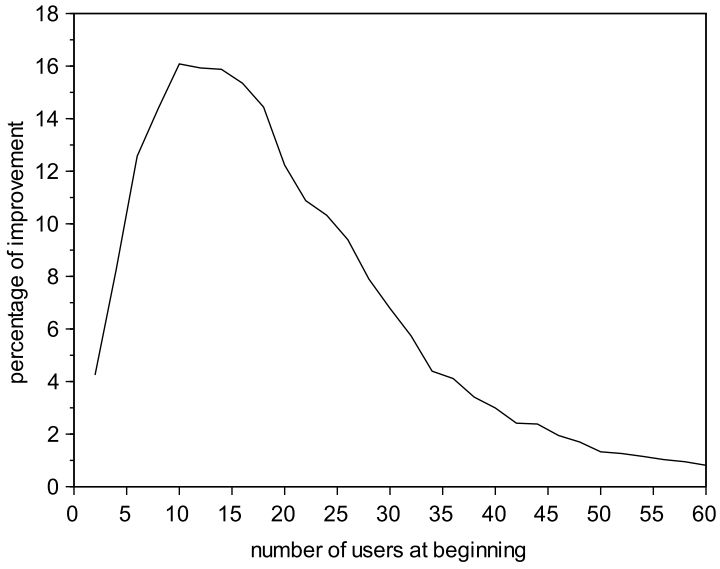

Figure 4: Percentage of improvement of WI policy over greedy in case of no arrival and one class of users.

values of K, both WIP and Greedy have the same performance but for moderate number of users one can gain up to 17% with WIP. The percentage improvement of the Whittle policy over the greedy policy is shown in Fig. 4 as a function of the initial number of cars in the system. The best performance improvement is obtained in moderate loads (initial number of users divided by the number of slots). When the load is high, the probability of having some user in a slot with high rate (close to the AP) is high. Thus, both greedy and Whittle policies choose similar users to serve and have similar performance.

Fig. 5 shows the number of users that were able of send all their data before leaving the coverage range. The x axis the the time-slot. Since there are 100 spatial-slots and users move one spatial-slot in one time-slot, at the end of 100 time-slots, the system is guaranteed to be empty and the simulation can stop. The initial number of cars for this set of experiments is $K = 20$. Here too, the Whittle performs better than the other policies.

4.2 New arrivals

We now evaluate the performance of the policies when there are new arrivals to the systems. Recall that in each time-slot a new user arrives with a probability p in the left-most spatial-slot. We first compare the policies for a small number of spatial-slots, $N = 11$. This allows use to compute the average reward of the optimal policy. In Fig. 6, the average total reward is plotted for the policies. Since the performance of the Gittin's index is similar to the greedy in the no arrival case, we omit this in the plot. We observe that Whittle policy almost overlaps with the optimal policy, and outperforms the others.

As a final comparison, we show the performance of the different polices (except the optimal one) for $N = 100$ spatial-slots and three classes of users. The values of the mean service requirement of the three classes are: $\eta_1^{-1} = 0.8$, $\eta_2^{-1} = 1.4$, and $\eta_3^{-1} = 4.2$. This time the optimal policy is not shown because the state-space is too big to allow for its computation. Fig. 7 shows the average total reward as a function of the probability of new arrival. It is observed that the Whittle policy performs much better than the greedy policy when there are more number of classes. The improvement is visible for a larger range of the probability of new

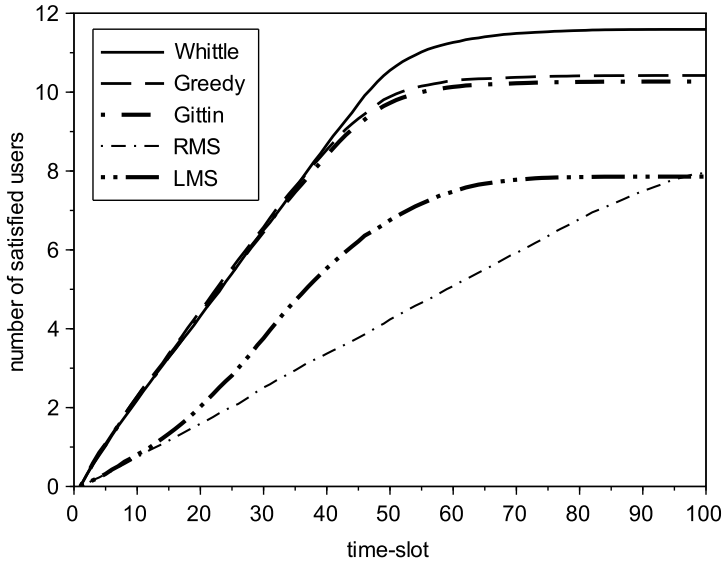

Figure 5: Number of satisfactory cars of five policies in the case of no arrival when there are 20 users at the beginning.

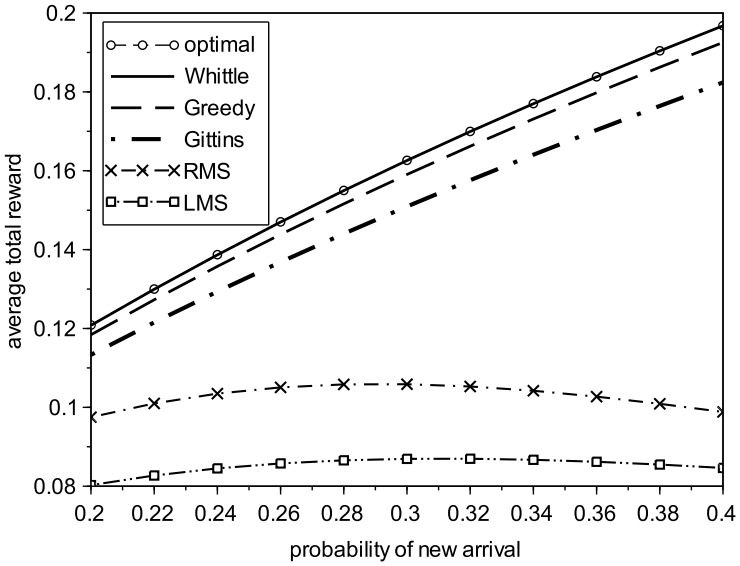

Figure 6: Comparison of policies when there are new arrivals and number of spatial-slots is small ($N = 11$) and one class of users.

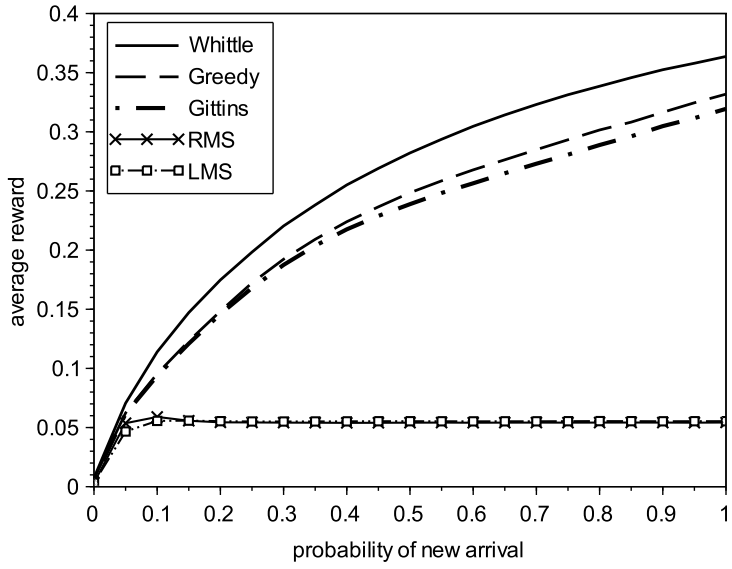

Figure 7: Comparison of policies when there are new arrivals and number of spatial-slots is large ($N = 100$) and three classes of users with mean service $\eta = 1$.

arrivals compared to when there is only class of users.

5 Conclusions and future work

We proposed a heuristic policy based on Whittle's restless multi-armed bandit framework for scheduling users in a drive-thru internet scenario. The simplified problem of no-arrivals has a nice structure, inherited from the mobility model of the drive-thru internet, which permits for simple computation of Whittle's indices. Between two users who have the same rate but who are on the opposite sides on the access point, the Whittle-index policy gives priority to the user on the right because the user on the left can be served later on. It was seen from numerical experiments that the heuristic policy based on Whittle's indices outperforms the greedy policy in various settings including dynamic arrivals and heterogeneous users.

This framework opens several interesting questions for related to the suboptimality of the proposed heuristics as well as generalizations of indexability to models with users moving on larger networks and with varying speeds.

Appendix

5.1 Proof of Theorem 1

We divide the proof of Theorem 1 into three steps:

- Proof of the existence of the solution of (5), which is given in Proposition 3,
- Proof of Theorem 1.1 for the states on the right. This is presented in Proposition 4,
- Proof of Theorem 1.2 for the states on the left. This is presented in Proposition 5.

Proposition 3. *For each x, the* (5) *has at least one solution $\nu \in (0, +\infty)$. The same holds for* (6).

For the proof of Prop. 3, we need the following two lemmas.

Lemma 2. *If $\nu = 0$, $V^1(x,0) > V^0(x,0)$, for $x = 0, 1, \ldots N$.*

Proof. We prove by induction in reverse direction, i.e, from state N to state 0. For $x = N$:

$$V^1(N,0) = r_N\eta > 0 = V^0(N,\nu),$$

so that $V(N,0) = V^1(N,0)$.

Suppose the claim is true until state x, i.e, $V^1(y,0) > V^0(y,0)$ for any $y \geq x$ and $V(y,0) = V^1(y,0)$. Then, by value iteration we have, for any $y \geq x$,

$$\begin{aligned} V(y,0) &= V^1(y,0) = r_y\eta + (1 - r_y\eta)V(y+1,0) \\ &= \ldots \\ &= 1 - (1 - r_y\eta)(1 - r_{y+1}\eta)\cdots(1 - r_N\eta). \end{aligned}$$

We now prove that the claim is true for state $x - 1$. We have:

$$V^0(x-1,0) = V(x,0) = 1 - (1 - r_x\eta)(1 - r_{x+1}\eta)\cdots(1 - r_N\eta)$$

and

$$\begin{aligned} V^1(x-1,0) &= r_{x-1}\eta + (1 - r_{x-1}\eta)V(x,0) \\ &= 1 - (1 - r_{x-1}\eta)(1 - r_x\eta)\cdots(1 - r_N\eta). \end{aligned}$$

Therefore $V^1(x-1,0) > V^0(x-1,0)$. □

Lemma 3. *For any $\nu > \max_{x=0,1,\ldots N}\{r_x\eta\}$, $V^1(x,\nu) < V^0(x,\nu)$, $x = 0, 1, \ldots N$.*

Proof. We prove by induction in reverse direction from state N to state 0. For $x = N$,

$$V^1(N,0) = r_N\eta - \nu < 0 = V^0(N,0),$$

since $r_x\eta < \nu$. So, $V(N,0) = V^1(N,0)$.

Suppose the claim is true until state x, i.e, $V^1(y,0) < V^0(y,0)$ so that, for any $y \geq x$, $V(y,0) = V^0(y,0)$. Then, by value interaction we have:

$$V(y,0) = 0$$

for any $y \geq x$. Thus, for state $x - 1$, we have

$$V^1(x-1,0) = r_{x-1}\eta - \nu < 0 = V^0(x-1,0).$$

□

Proof of Prop.3. From Lemma 2, we know that $V^1(x,0) > V^0(x,0)$. From Lemma 3, we have $V^1(x,1) > V^0(x,1)$ (we can take $\nu = 1$ because $r_x\eta$ is a probability and is thus smaller than 1). Using these two lemmas and the continuity of the function $V^1(x,\cdot), V^0(x,\cdot)$ in the second variable, we can conclude that (5) each has at least one solution. □

We now prove the uniqueness of the solution of (5) on the right hand side, and give its properties. We also characterize the behaviour of the value function which will be used later for the proof of Theorem 1.

Proposition 4. *(For Theorem 1.1) For any $x \geq N/2$, Eqn. (6) has a unique solution ν_x. Moreover, we have:*

1. *$\nu_x = r_x\eta$. Thus, $\nu_{N/2} > \nu_{N/2+1} > \cdots > \nu_N$.*

2. *The value function takes the following form:*
   ⋆ *If $\nu \geq r_x\eta$, then*

$$V(x,\nu) = 0, \tag{8}$$

   ⋆ *If $0 \leq \nu < r_x\eta$, then*

$$V(x,\nu) = \sum_{i=x}^{y} \left(\prod_{j=x}^{i-1} (1 - r_j\eta) \right) (r_i\eta - \nu), \tag{9}$$

   where $y \in \{x, x+1, ..., N\}$ is such that $r_{y+1}\eta \leq \nu < r_y\eta$, with the convention $r_{N+1} = 0$.

For the proof of this proposition, we need the following lemmas which are proven in part 5.5 - the additional proofs.

Lemma 4. *If $A_1, B_1, A_2, B_2, ..., A_k, B_k > 0, k \geq 2$ and $\frac{A_1}{B_1} < \frac{A_2}{B_2} < \cdots < \frac{A_k}{B_k}$ then*

$$\frac{A_1}{B_1} < \frac{A_1 + A_2}{B_1 + B_2} < \cdots < \frac{A_1 + A_2 + \cdots + A_k}{B_1 + B_2 + \cdots + B_k}.$$

Lemma 5. *If $A_1 > A_2 + A_3 + \cdots + A_k > 0, B_1 > B_2 + B_3 + \cdots + B_k > 0, k \geq 2$ and $\frac{A_1}{B_1} > \frac{A_2}{B_2} > \cdots > \frac{A_k}{B_k}$ then*

$$\frac{A_1}{B_1} < \frac{A_1 - A_2}{B_1 - B_2} < \cdots < \frac{A_1 - A_2 - \cdots - A_k}{B_1 - B_2 - \cdots - B_k}.$$

For every $\Delta > x$, we define:

$$a(x,\Delta) := 1 - \sum_{i=x+1}^{\Delta} r_i\eta \prod_{j=x+1}^{i-1} (1 - r_j\eta),$$

$$b(x,\Delta) := \sum_{i=x+1}^{\Delta} \prod_{j=x+1}^{i-1} (1 - r_j\eta).$$

Proof of Prop. 4. For the states on the right side ($x \geq N/2$), we prove by induction.
We start by state N. From the boundary condition, $V(N+1,\nu) = 0$. so that

$$V(N,\nu) = \max\{r_N\eta - \nu, 0\}.$$

By Definition 2, ν_N is such that there is no difference between being active and being passive in state N. Thus, $r_N\eta - \nu_N = 0$, and so $\nu_N = r_N\eta$.

We assume that the claim is true until state $x+1$, i.e, for any $y = x+1, x+2, ..., N$ we have $\nu_y = r_y\eta$ and formula of (9) for the value function.

Now we show the claim holds for state x. Assume by contradiction that $\nu_x \leq \nu_{x+1}$ so there exists $y \geq x+1$ such that $\nu_x \in [r_{y+1}\eta, r_y\eta)$. Using the induction hypotheses, the solution of equation (6) is obtained as:

$$\nu_x = \frac{r_x\eta a(x,y)}{1 - r_x\eta b(x,y)}.$$

For $k = 1, \ldots, y - x$, define:

$$\begin{aligned} A_k &= r_x\eta \cdot r_{x+k}\eta \prod_{j=x+1}^{x+k-1} (1 - r_j\eta). \\ B_k &= r_x\eta \cdot \prod_{j=x+1}^{x+k-1} (1 - r_j\eta), \end{aligned}$$

Applying Lemma 5 with the observation that:

$$\frac{A_1}{B_1} = r_{x+1}\eta > r_{x+2}\eta = \frac{A_2}{B_2} > \cdots > \frac{A_{y-x}}{B_{y-x}} = r_y\eta,$$

and

$$\frac{A_1 - A_2 - \cdots - A_{y-x}}{B_1 - B_2 - \cdots - B_{y-x}} = \nu_x,$$

we get $r_{x+1}\eta = \nu_{x+1} \geq \nu_x = \frac{A_1 - A_2 - \cdots - A_{y-x}}{B_1 - B_2 - \cdots - B_{y-x}} > \frac{A_1}{B_1} = r_{x+1}\eta$, which is a contradiction.

Therefore, $\nu_x > \nu_{x+1}$. Using (5) for state x and (8) for states $x+1, ..., N$, we obtain $\nu_x = r_x\eta$. The formulas (8) and (9) for state x follow by using induction on Bellman's equation (4). □

Next, we move to the proof of Theorem 1.2.

Proposition 5. *(For Theorem 1.2) For every state on LHS $x \leq N/2 - 1$, Eqn. (6) has a unique solution ν_x. Moreover, we have:*

1. $$\nu_x = f(x, \Delta(x)),$$
   where $\Delta(x) \geq N/2$ is chosen such that $f(x, \Delta(x)) \in [r_{\Delta(x)+1}\eta, r_{\Delta(x)}\eta)$.
2. *ν_x increases in x on LHS, i.e,*
   $$\nu_0 < \nu_1 < \nu_2 < \cdots < \nu_{N/2-1} < r_{N/2}\eta = \nu_{N/2}.$$
3. *The value function has the following form:*
   - ⋆ *If $\nu \geq \nu_x$, then $V(x, \nu) = V(x + 1, \nu)$.*
   - ⋆ *If $0 \leq \nu < \nu_x$, then*
   $$V(x, \nu) = (r_x\eta - \nu) + (1 - r_x\eta)V(x + 1, \nu).$$

Before proving Proposition 5, we need to characterize properties of function $f(x, \Delta)$ which are described in the following Lemma 6. Remark that we can rewrite f as:

$$f(x, \Delta) = \frac{r_x\eta a(x, \Delta)}{1 - r_x\eta b(x, \Delta)}.$$

Lemma 6. *For a fixed $x \in L = \{0, 1, ..., N/2 - 1\}$, define $D_x = \{\Delta | \Delta \geq N/2, f(x, \Delta) \geq 0\}$. Recall from (7) that f is defined only on integers. Let $\Delta(x) \in D_x$ be the smallest value for which $r_{\Delta(x+1)}\eta \leq f(x, \Delta(x)) < r_{\Delta(x)}\eta$. Then,*

$$\Delta(x) = \arg \min_{\Delta \in D_x} f(x, \Delta).$$

Moreover, for fixed x and considering $f(x, \Delta)$ as a function of Δ in D_x, then f decreases from $N/2$ to $\Delta(x)$ and increases from $\Delta(x)$ to b. Outside of D_x, $f(x, \Delta)$ is either negative or infinity.

Proof. First, we will show that D_x is connected. The numerator of $f(x,\Delta)$ is $r_x\eta \cdot a(x,\Delta)$ where

$$a(x,\Delta) = 1 - \sum_{i=x+1}^{\Delta} r_i\eta \prod_{j=x+1}^{i-1}(1-r_j\eta) = \prod_{j=x+1}^{\Delta}(1-r_j\eta) \geq 0.$$

So, the numerator of $f(x,\Delta)$ is always positive.

Since the numerator of $f(x,\Delta)$ is always positive, the sign of f depends on the sign of its denominator. It is easy to see that the denominator of f is decreasing in Δ because b is an increasing function of Δ. So, for $\Delta_2 > \Delta_1 \geq N/2$, if $f(x,\Delta_2)$ is positive, then so is $f(x,\Delta_1)$. This implies that D_x is connected. And, for $\Delta \in \{b+1, b+2, ..., N\}$, $f(x,\Delta)$ is either negative or infinity.

Now, we will prove the claim in the lemma. Suppose $\Delta_1 \in D_x = \{N/2, N/2+1, ..., b\}$ for some b (b depends on x), and $\Delta_2 > \Delta_1 > \Delta(x)$. Define

$$\begin{aligned} A_1 &= r_x\eta \cdot a(x,\Delta(x)), \\ B_1 &= 1 - r_x\eta \cdot b(x,\Delta(x)), \end{aligned}$$

and, for $2 \leq k \leq \Delta_2 - \Delta(x) + 1$,

$$\begin{aligned} A_k &= r_x\eta \cdot r_{\Delta(x)+k-1}\eta \prod_{j=x+1}^{\Delta(x)+k-2}(1-r_j\eta) \\ B_k &= r_x\eta \cdot \prod_{j=x+1}^{\Delta(x)+k-2}(1-r_j\eta). \end{aligned}$$

We have $B_1 - (B_2 + \cdots + B_{\Delta_2-\Delta(x)+1}) > 0$ since it is the denominator of $f(x,\Delta_2)$ with $\Delta_2 \in D_x$. And, $A_1 - (A_2 + \cdots + A_{\Delta_2-\Delta(x)+1}) > 0$ since it is the numerator of $f(x,\Delta_2)$ which is positive. Moreover,

$$\frac{A_1}{B_1} = f(x,\Delta(x)) \in [r_{\Delta(x)+1}\eta, r_{\Delta(x)}\eta),$$

and we get

$$\frac{A_1}{B_1} \geq r_{\Delta(x)+1}\eta = \frac{A_2}{B_2} > \cdots > r_{\Delta_2}\eta = \frac{A_{\Delta_2-\Delta(x)+1}}{B_{\Delta_2-\Delta(x)+1}}.$$

Applying Lemma 5 we get:

$$\begin{aligned} f(x,\Delta_1) &= \frac{A_1 - A_2 - \cdots - A_{\Delta_1-\Delta(x)+1}}{B_1 - B_2 - \cdots - B_{\Delta_1-\Delta(x)+1}} \\ &< \frac{A_1 - A_2 - \cdots - A_{\Delta_2-\Delta(x)+1}}{B_1 - B_2 - \cdots - B_{\Delta_2-\Delta(x)+1}} = f(x,\Delta_2). \end{aligned}$$

Hence, if $\Delta_2 > \Delta_1 > \Delta(x)$ and $\Delta_1, \Delta_2 \in D_x = \{\Delta(x), \Delta(x)+1, ..., b\}$ then $f(x,\Delta_2) > f(x,\Delta_1)$.Similarly, by using Lemma 4, if $N/2 \leq \Delta_1 < \Delta_2 < \Delta(x)$ then $f(x,\Delta_1) > f(x,\Delta_2)$.

Combining the above two results, we get the minimum of function $f(x,\Delta)$ in D_x attained when $\Delta = \Delta(x)$. Further, if $\Delta \in \mathcal{N}^+ \setminus D_x$, then $f(x,\Delta)$ is either negative or infinity. □

The existence of $\Delta(x)$ will be proved in Lemma 7 for $x = N/2 - 1$ and for other values of x in the proof of Proposition 4 using the existence of the solution of (5).

Before doing induction for the states on the left hand side, we first consider state $N/2 - 1$.

Lemma 7. *$\Delta(N/2-1) \geq N/2$ exists. Further, the equation $V^1(N/2-1,\nu) = V^0(N/2-1,\nu)$ has the unique solution which is:*

$$\nu_{N/2-1} = f(N/2-1, \Delta(N/2-1)).$$

Moreover,
1. If $\nu < \nu_{N/2-1}$, then $V(N/2-1,\nu) = (r_{N/2-1}\eta - \nu) + (1 - r_{N/2-1}\eta)V(N/2,\nu)$,
2, If $\nu \geq \nu_{N/2-1}$, then $V(N/2-1,\nu) = V(N/2,\nu)$,
with $V(N/2,\nu)$ given in Prop. 4.

Proof. By Prop. 3, the equation

$$r_{N/2-1}\eta - \nu = r_{N/2-1}\eta V(N/2,\nu)$$

has at least one solution ν_1. If $\nu_1 \geq r_{N/2}\eta$ then the left-hand side of the above equation is negative while the right-hand side equals 0 (which is a impossible).

So, $\nu_1 < r_{N/2}\eta$. Then, there exists $\Delta_1 \geq N/2$ such that $\nu_1 \in [r_{\Delta_1+1}\eta, r_{\Delta_1}\eta)$. By developing $V(N/2,\nu_1)$ in Prop. 4,

$$r_{N/2-1}\eta - \nu_1 = r_{N/2-1}\eta \sum_{i=N/2}^{\Delta_1} \left(\prod_{j=x}^{i-1} (1 - r_j\eta) \right) (r_i\eta - \nu). \tag{10}$$

Solving this linear equation, we get:

$$\begin{aligned} \nu_1 &= f(N/2-1, \Delta_1) \\ &= \frac{r_{N/2-1}\eta\big(1 - \sum_{i=N/2}^{\Delta_1} r_i\eta \prod_{j=N/2}^{i-1}(1-r_j\eta)\big)}{1 - r_{N/2-1}\eta(\sum_{i=N/2}^{\Delta_1} \prod_{j=N/2}^{i-1}(1-r_j\eta))}. \end{aligned}$$

Obviously every solution is of the above form. Suppose Δ_1 is the first state from the left such that

$$f(x, \Delta_1) \in [r_{\Delta_1+1}\eta, r_{\Delta_1}\eta). \tag{*}$$

We now prove the uniqueness of the solution by contradiction. Suppose there exist another solution $\nu_2 < \nu_1$ (ν_2 can not greater than ν_1 due to (*). We know that ν_2 can not be in $[r_{\Delta_1+1}\eta, r_{\Delta_1}\eta)$ since the linear equation (10) has a unique solution. Therefore, there exists $\Delta_2 > \Delta_1$ such that $\nu_2 \in [r_{\Delta_2+1}\eta, r_{\Delta_2}\eta)$. Following the same approach as for finding ν_1, we get:

$$\nu_2 = f(N/2-1, \Delta_2).$$

We have 2 cases:

- If $\Delta_2 \in D_x$ then $f(x, \Delta_2) > 0$. By Lemma 6 we have:

  $$\nu_2 = f(N/2-1, \Delta_2) \geq f(N/2-1, \Delta_1) = \nu_1$$

  This contradicts $\nu_2 < \nu_1$.

- If $\Delta_2 \geq N/2$ but is not in D_x then $\nu_2 = f(N/2-1, \Delta_2)$ is negative or infinity which cannot be true since we solve the equation in $[0, +\infty)$.

Thus, there is a unique Δ_1 if it exists.

The existence of $\Delta_1 \geq N/2$ such that $f(N/2-1, \Delta_1) \in [r_{\Delta_1+1}\eta, r_{\Delta_1}\eta)$ follows from the fact that the value function is piece-wise linear in Δ and that there exists a solution to (5).

Hence, the solution exists and is unique. Denote the solution by $\nu_{N/2-1}$. Let $\Delta(N/2-1) := \Delta_1$ and we get the relationship:

$$\begin{aligned}\nu_{N/2-1} &= f(N/2-1, \Delta(N/2-1)) \\ &\in [r_{\Delta(N/2-1)+1}\eta, r_{\Delta(N/2-1)}\eta).\end{aligned}$$

By the continuity of $V^1(x,\cdot), V^0(x,\cdot)$ in the second variable, the fact that $V^1(x,0) > V^0(x,0)$ (from Lemma 2), and $V^1(x,1) < V^0(x,1)$ (from Lemma 3) and the uniqueness of solution proved above, we get the conclusion on the value function. □

In a similar way, we will prove Prop. 5 by induction on states $x \leq N/2-1$. That is, we will show that (5) has a unique solution, called ν_x, and, on the left-hand side ν_x increases in x, i.e if $x < y \leq N/2-1$ then $\nu_x < \nu_y$.

Proof of Proposition 5. We shall prove the claim by induction in the reverse direction. For state $N/2-1$, the claim follows from Lemma 7. Suppose the claim is true until state $x \leq N/2-1$. We now prove for state $x-1$. Consider the equation:

$$r_{x-1}\eta - \nu = r_{x-1}\eta V(x,\nu).$$

By Lem. 3 we know that there is at least one solution. Suppose ν is a solution of this equation.

- If $\nu \geq r_{N/2}\eta$, then

  $$r_{x-1}\eta - \nu = r_{x-1}\eta V(x,\nu) = \cdots = r_{x-1}\eta V(N+1) = 0,$$

  which follows by the induction hypothesis for all states in $[x, x+1, ..., N/2-1]$ and by Proposition 4, for all states in $[N/2, N/2+1, ..., N]$. This implies that $\nu = r_{x-1}\eta < r_{N/2}\eta$, which leads to a contradiction with $\nu \geq r_{N/2}\eta$.

- If $\nu_x \leq \nu < \nu_{N/2}$ then there exist y_1, y_2 such that:

  $$\begin{cases} x \leq y_1 \leq N/2-1, N/2 \leq y_2 \leq \Delta(y_1) \\ \nu \in [\nu_{y_1}, \nu_{y_1+1}) \cap [r_{y_2+1}\eta, r_{y_2}\eta). \end{cases} \tag{11}$$

  So, by induction hypothesis and Prop. 4, we can develop $V(x,\nu)$ to get:

  $$\begin{aligned} V(x,\nu) &= V(x+1,\nu) = \cdots = V(y_1+1,\nu) \\ &= (r_{y_1+1}\eta - \nu) + (1 - r_{y_1+1})V(y_1+2,\nu) \\ &= \cdots \\ &= \sum_{i=y_1+1}^{y_2} \left(\prod_{j=y_1+1}^{i-1} (1 - r_j\eta) \right) (r_i\eta - \nu). \end{aligned}$$

  Now, the equation $r_{x-1}\eta - \nu = r_{x-1}\eta V(x,\nu)$ becomes linear in ν, which can be solved to get:

  $$\nu = \frac{r_{x-1}\eta a(y_1,y_2)}{1 - r_{x-1}\eta b(y_1,y_2)}. \tag{12}$$

  We have:

  $$\nu = \frac{r_{x-1}\eta a(y_1,y_2)}{1 - r_{x-1}\eta b(y_1,y_2)} < \frac{r_{y_1}\eta a(y_1,y_2)}{1 - r_{y_1}\eta b(y_1,y_2)} = f(y_1,y_2), \tag{13}$$

  and we remark that $y_2 \leq \Delta(y_1)$.

  Now, there are two sub-cases:

– If $y_2 = \Delta(y_1)$ then

$$\nu_{y_1} = f(y_1, \Delta(y_1)) = f(y_1, y_2) > \nu,$$

where the last inequality is due to (13). This contradicts $\nu \geq \nu_{y_1}$ in (11).

– Suppose $y_2 < \Delta(y_1)$. From (11), we have $\nu \in [r_{y_2+1}\eta, r_{y_2}\eta)$, and by (13) we have $\nu < f(y_1, y_2)$. Therefore, $f(y_1, y_2) > r_{y_2+1}\eta$. Note that the statement $f(y_1, y_2) < r_{y_2}\eta$ cannot be true because this will imply that $\Delta(y_1) = y_2$. On the other hand, by induction hypothesis, $\Delta(y_1)$ is the unique state that satisfies $f(y_1, \Delta(y_1)) \in [r_{\Delta(y_1)+1}\eta, r_{\Delta(y_1)}\eta)$ and we have assumed that $y_2 < \Delta(y_1)$. Thus,

$$f(y_1, y_2) \geq r_{y_2}\eta. \tag{14}$$

Define

$$\begin{aligned} A_1 &= r_{y_1}\eta \cdot a(y_1, y_2), \\ B_1 &= 1 - r_{y_1}\eta \cdot b(y_1, y_2), \end{aligned}$$

and for $k = 2, \ldots, \Delta(y_1) - y_2 + 1$,

$$\begin{aligned} B_k &= r_{y_1}\eta \prod_{j=y_1+1}^{y_2+k-2} (1 - r_j\eta), \\ A_k &= r_{y_2+k-1} \cdot B_k, \end{aligned}$$

From the above definitions and (14),

$$\begin{aligned} \frac{A_1}{B_1} &= f(y_1, y_2) \geq r_{y_2}\eta > \frac{A_2}{B_2} = r_{y_2+1}\eta > \cdots \\ &> \frac{A_{\Delta(y_1)-y_2+1}}{B_{\Delta(y_1)-y_2+1}} = r_{\Delta(y_1)}\eta. \end{aligned}$$

Apply Lemma 5 on A_k and B_k, we get:

$$\begin{aligned} r_{y_2}\eta &\leq f(y_1, y_2) = \frac{A_1}{B_1} \\ &< \frac{A_1 - A_2 - \cdots A_{\Delta(y_1)-y_2+1}}{B_1 - B_2 - \cdots - B_{\Delta(y_1)-y_2+1}} \\ &= f(y_1, \Delta(y_1)) < r_{\Delta(y_1)}\eta, \end{aligned}$$

which is in contradiction to $N/2 \leq y_2 < \Delta(y_1)$.

- Finally, suppose $0 < \nu < \nu_x$. Then, following similar arguments as in the proof of Lemma 7 and by existence of the solution of (6), there exists a unique $\Delta(x-1) \geq \Delta(x)$ such that

$$\nu = f(x-1, \Delta(x-1)) \in [r_{\Delta(x-1)+1}\eta, r_{\Delta(x)}\eta).$$

We have proved the first two claims of Proposition 5.

Now, we prove the third claim. From Lemma 2, we have $V^1(x, 0) > V^0(x, 0)$ and from Lemma 3, we know that $V^1(x, \infty) < V^0(x, \infty)$. Since ν_x is the unique solution of $V^1(x, 0) = V^0(x, 0)$ and $V^1(x, \nu)$ and $V^0(x, \nu)$ are continuous in ν, we can infer that $V^1(x, \nu) \geq V^0(x, \nu)$ in $[0, \nu_x]$ and $V^1(x, \nu) \leq V^0(x, \nu)$ in $[\nu_x, \infty)$. This implies the claimed form of the value function for state x. □

5.2 Proof of Proposition 1

Proof. We prove it by induction for the direction from $N/2-1$ to 0 by showing that

$$\nu_x < \nu_{N-x} = r_{N-x}\eta = r_x\eta \text{ for any } x \leq N/2-1. \tag{15}$$

Firstly, we prove for state $x = N/2-1$. Suppose that $\nu_{N/2-1} \geq \nu_{N/2+1} = r_{N/2+1}\eta = r_{N/2-1}\eta$. By proof of Lemma 7 we have $\nu_{N/2-1} < r_{N/2}\eta$. It implies that $\nu_{N/2-1} \in [r_{N/2+1}\eta, r_{N/2}\eta)$, so $\Delta(N/2-1) = N/2$. Therefore,

$$\begin{aligned} r_{N/2-1}\eta &\leq \nu_{N/2-1} \\ &= f(N/2-1, \Delta(N/2-1)) \\ &= \frac{r_{N/2-1}\eta(1-r_{N/2}\eta)}{1-r_{N/2-1}\eta}. \end{aligned}$$

But $r_{N/2}\eta > r_{N/2-1}\eta$, so $r_{N/2-1}\eta\frac{(1-r_{N/2}\eta)}{1-r_{N/2-1}\eta} \leq r_{N/2-1}\eta$. This is a contradiction.

Next we prove the claim by induction for $x < N/2-1$. Let (15) be true until state $x+1$. Suppose it is not true for x, then $f(x,\Delta(x)) \geq r_x\eta$. By induction, we get $f(x+1,\Delta(x+1)) < r_{x+1}\eta$. By theorem 1 we have $f(x+1,\Delta(x+1)) > f(x,\Delta(x))$. Therefore $r_x\eta \leq f(x,\Delta(x)) < f(x+1,\Delta(x+1)) < r_{x+1}\eta$, it implies $\Delta(x) = \Delta(x+1)$ and $\nu_{x+1} = f(x+1,\Delta(x))$. To get the contradiction we shall show that :

$$\nu_{x+1} - \nu_x = f(x+1,\Delta(x)) - f(x,\Delta(x)) > r_{x+1}\eta - r_x\eta.$$

Indeed,

$$\begin{aligned} \nu_{x+1} - \nu_x = &\ \frac{r_{x+1}\eta a(x+1,\Delta(x))}{1-r_{x+1}\eta b(x+1,\Delta(x))} - \frac{r_x\eta a(x,\Delta(x))}{1-r_x\eta b(x,\Delta(x))} \\ > &\ \frac{r_{x+1}\eta a(x,\Delta(x))}{1-r_{x+1}\eta b(x,\Delta(x))} - \frac{r_x\eta a(x,\Delta(x))}{1-r_x\eta b(x,\Delta(x))} \qquad (16) \\ = &\ \frac{\big(r_{x+1}\eta - r_x\eta\big)a(x,\Delta(x))}{\big(1-r_{x+1}\eta b(x,\Delta(x))\big)\big(1-r_x\eta b(x,\Delta(x))\big)} \\ > &\ r_{x+1}\eta - r_x\eta \text{ (contradiction).} \qquad (17) \end{aligned}$$

The inequality (16) is implied by using Lemma 5 for $\frac{A1}{B1} = \frac{r_{x+1}\eta a(x,\Delta(x))}{1-r_{x+1}\eta b(x,\Delta(x))}$, $\frac{A2}{B2} = \frac{r_{x+1}\eta * r_{x+1}\eta}{r_{x+1}\eta}$ and $\frac{A1-A2}{B1-B2} = \frac{r_{x+1}\eta a(x+1,\Delta(x))}{1-r_{x+1}\eta b(x+1,\Delta(x))}$.
The inequality (17) is due to $f(x,\Delta(x)) = \frac{r_x\eta a(x,\Delta(x))}{1-r_x\eta b(x,\Delta(x))} \geq r_x\eta$, so $\frac{a(x,\Delta(x))}{1-r_x\eta b(x,\Delta(x))} \geq 1$, and $\frac{a(x,\Delta(x))}{\big(1-r_{x+1}\eta b(x,\Delta(x))\big)\big(1-r_x\eta b(x,\Delta(x))\big)} > 1$. □

5.3 Proof of Proposition 2

Proof. Suppose that car of type 2 leaves after state $N_1 < N$. Remark that in this case the rate of the type-2 car equals to 0 after state N_1.

- If x is on the right hand side, by Proposition 4 we get $\nu_x^1 = \nu_x^2 = r_x\eta$.
- If x is on the left hand side, we show that $\Delta^1(x) \leq \Delta^2(x)$ where $\Delta^1(x), \Delta^2(x)$ are given in proposition 5 for type 1, and type 2 respectively.
  - If $\Delta^2(x) \leq N_1 - 1$ then $\Delta^2(x) = \Delta^1(x)$ due to the rate function being same until state N_1.

– $\Delta^2(x) = N_1$ we consider 2 sub-cases:
⋆ $f(x, \Delta^2(x)) \in [r_{\Delta^2(x)+1}\eta, r_{\Delta^2(x)}\eta)$, it implies $\Delta^2(x) = \Delta^1(x)$ due to same rate function until state N_1.
⋆ $f(x, \Delta^2(x)) \in [0, r_{\Delta^2(x)+1}\eta)$, it implies that $\Delta^1(x) > \Delta^2(x)$. Denote D_x^1 (resp. D_x^2) the domain of x for type-1 (resp. type-2) car. Then $D_x^2 \subset D_x^1$ since $\Delta^1(x) > \Delta^2(x)$. By Lemma 6, we get $f(x, \Delta^1(x))$ attains minimum at $\Delta^1(x)$ in the domain D_x^1 so $f(x, \Delta^1(x)) \leq f(x, \Delta^2(x))$.

5.4 Computation of the Gittins' index

Gittins' index policy is known optimal in the case of static bandit, but it not for the bandits are restless. By static bandit, we mean a bandit who does not change state when it is not activated. In our problem, the users change state even if they are not activated (or served). Thus, the users are restless and do not fall in the static framework of Gittins. However, for comparison purposes, we also include the Gittins' index.

Gittin's index for state x is defined as the follows:

$$G_k(x) = \sup_{\Delta \in \{1,2,...,N-x+1\}} \frac{\sum_{t=0}^{\tau_\Delta - 1} \beta^t \mathbb{E}(R_k(X_k(t))|X_k(0) = x)}{\sum_{t=0}^{\tau_\Delta - 1} \beta^t}, \tag{18}$$

where $\hat{\tau} = \inf\{t \geq 0 : X_k(t) = N + 1\}$ and $\tau_\Delta = \min(\Delta, \hat{\tau})$ and here we take $\beta = 1$.
To compute $G_k(x)$ we define:

$$G_k^\Delta(x) = \frac{\sum_{t=0}^{\tau_\Delta - 1} \beta^t \mathbb{E}(R_k(X_k(t))|X_k(0) = x)}{\sum_{t=0}^{\tau_\Delta - 1} \beta^t},$$

for all $\Delta \in \{1, 2, ..., N - x + 1\}$. We have:

- $\mathbb{P}(\tau_\Delta = i|X_k(0) = x) = \mathbb{P}(\hat{\tau} = i|X_k(0) = x) = \mathbb{P}(X_k(i) = N + 1|X_k(0) = x, X_k(1) \neq N + 1, ..., X_k(i - 1) \neq N + 1) = (1 - r_x\eta_k)(1 - r_{x+1}\eta_k)...(1 - r_{x+i-2}\eta_k) * r_{x+i-1}\eta_k$, for $i = 1, 2, ..., \Delta - 1$
- $\mathbb{P}(\tau_\Delta = \Delta|X_k(0) = x) = \mathbb{P}(\hat{\tau} \geq \Delta|X_k(0) = x) = \mathbb{P}(X_k(0) = x, X_k(1) \neq N+1, ..., X_k(\Delta-1) \neq N + 1) = (1 - r_x\eta_k)(1 - r_{x+1}\eta_k)...(1 - r_{x+\Delta-2}\eta_k)$.

So $G_k^\Delta(x)$ equals to:

$$\frac{\sum_{i=1}^{\Delta} \mathbb{P}(\tau_\Delta = i) \sum_{t=0}^{i-1} \beta^t \mathbb{E}(R_k(X_k(t))|X_k(0) = x, \tau_\Delta = i)}{\sum_{i=0}^{\Delta} \mathbb{P}(\tau_\Delta = i) \sum_{t=0}^{i-1} \beta^t} = \frac{\sum_{i=1}^{\Delta} \mathbb{P}(\tau_\Delta = i) \sum_{t=0}^{i-1} r_{x+t}\eta_k}{\sum_{i=0}^{\Delta} \mathbb{P}(\tau_\Delta = i) * i},$$

We can find $\sup_{\Delta \in \{1,2,...,N-x+1\}} G_k^\Delta(x)$ numerically.

□

5.5 Additional proofs

Proof. (Lemma 4) We can prove directly that

$$\frac{A_1}{B_1} < \frac{A_1 + A_2}{B_1 + B_2} < \frac{A_2}{B_2},$$

But $\frac{A_2}{B_2} < \frac{A_3}{B_3}$ by the above assumption, it implies $\frac{A_1+A_2}{B_1+B_2} < \frac{A_3}{B_3}$. Therefore we have:

$$\frac{A_1 + A_2}{B_1 + B_2} < \frac{(A_1 + A_2) + A_3}{(B_1 + B_2) + B_3} < \frac{A_3}{B_3}.$$

By induction in that way we get the conclusion. □

Proof. (Lemma 5) The proof of this lemma is similar to the proof of Lemma 4 with the below observation:

$$\frac{A_k}{B_k} < \cdots \frac{A_3}{B_3} < \frac{A_2}{B_2} < \frac{A_1}{B_1} < \frac{A_1 - A_2}{B_1 - B_2}.$$

□

References

[1] Nan Cheng, Ning Lu, Ning Zhang, Xuemin (Sherman) Shen, and Jon W. Mark. Vehicular WiFi offloading. *Veh. Commun.*, 1(1):13–21, January 2014.

[2] Haibo Zhou, Lin Gui, Quan Yu, and Xuemin (Sherman) Shen. *Overview of Vehicular Communications in Drive-thru Internet*, pages 11–17. Springer International Publishing, Cham, 2015.

[3] D. Jia, K. Lu, J. Wang, X. Zhang, and X. Shen. A survey on platoon-based vehicular cyber-physical systems. *IEEE Communications Surveys Tutorials*, 18(1):263–284, Firstquarter 2016.

[4] J. Ott and D. Kutscher. Drive-thru internet: IEEE 802.11b for "automobile" users. In *IEEE INFOCOM 2004*, volume 1, page 373, March 2004.

[5] J. J. Alcaraz, J. Vales-Alonso, and J. Garcia-Haro. Link-layer scheduling in vehicle to infrastructure networks: An optimal control approach. *IEEE Journal on Selected Areas in Communications*, 29(1):103–112, January 2011.

[6] Qiang Zheng, Kan Zheng, Periklis Chatzimisios, and Fei Liu. Joint optimization of link scheduling and resource allocation in cooperative vehicular networks. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):170, Jun 2015.

[7] D. Jia, R. Zhang, K. Lu, J. Wang, Z. Bi, and J. Lei. Improving the uplink performance of drive-thru internet via platoon-based cooperative retransmission. *IEEE Transactions on Vehicular Technology*, 63(9):4536–4545, Nov 2014.

[8] Q. Wang, P. Fan, and K. B. Letaief. On the joint V2I and V2V scheduling for cooperative vanets with network coding. *IEEE Transactions on Vehicular Technology*, 61(1):62–73, Jan 2012.

[9] Richard R. Weber and Gideon Weiss. On an index policy for restless bandits. *Journal of Applied Probability*, 27(3):637–648, 1990.

[10] M. Larrnaaga, U. Ayesta, and I. M. Verloop. Dynamic control of birth-and-death restless bandits: Application to resource-allocation problems. *IEEE/ACM Transactions on Networking*, 24(6):3812–3825, December 2017.

[11] Dimitri P. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1987.

[12] P. Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 25:287–298, 1988.

[13] John Gittins, Kevin Glazebrook, and Richard Weber. *Multi-armed Bandit Allocation Indices*. John Wiley & Sons, 2011.

[14] Vivek S. Borkar and Sarath Pattathil. Whittle indexability in egalitarian processor sharing systems. *Annals of Operations Research*, 2017.

[15] Arjun Anand and Gustavo de Veciana. A Whittle's index based approach for qoe optimization in wireless networks. In *Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '18, pages 39–39, New York, NY, USA, 2018. ACM.

[16] C. Buyukkoc, P. Varaya, and J. Walrand. The $c\mu$ rule revisited. *Adv. Appl. Prob.*, 17:237–238, 1985.

[17] L.E. Schrage and L.W. Miller. The queue M/G/1 with the shortest remaining processing time discipline. *Operations Research*, 14:670–684, 1966.

[18] J.C. Gittins, K. Glazebrook, and R. Weber. *Multi-armed Bandit Allocation Indices*. Wiley, 2011.

[19] P. Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 25:287–298, 1988.

[20] C.H. Papadimitriou and J.N. Tsitsiklis. The complexity of optimal queueing network. *Mathematics of Operations Research*, 24(2):293–305, 1999.

[21] J. Niño-Mora. Dynamic priority allocation via restless bandit marginal productivity indices. *TOP*, 15(2):161–198, 2007.

[22] I.M. Verloop. Asymptotically optimal priority policies for indexable and non-indexable restless bandits. *Annals of Applied Probability*, 2016.

[23] Samuli Aalto, Pasi Lassila, and Prajwal Osti. Whittle index approach to size-aware scheduling with time-varying channels. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '15, pages 57–69, New York, NY, USA, 2015. ACM.

[24] Dimitri P. Bertsekas and John N. Tsitsiklis. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595, 1991.

[25] K.D. Glazebrook, C. Kirkbride, and J. Ouenniche. Index policies for the admission control and routing of impatient customers to heterogeneous servcie stations. *Operations Research*, 57:975–989, 2009.

[26] U. Ayesta, P. Jacko, and V. Novak. Scheduling of multi-class queueing system with abandonment. *Journal of Scheduling*, 20:129–145, 2017.

[27] P. S. Ansell, K. D. Glazebrook, J. Nino-Mora, and M. O'Keeffe. Whittle's index policy for a multi-class queueing system with convex holding costs. *Mathematical Methods of Operations Research*, 57(1):21–39, 2003.