Optimal non-preemptive scheduling with time-varying non-observable channels*

Thomas Hira*[1], Urtzi Ayesta[1], Rhonda Righter[2], and Ina Maria Verloop[1]

[1]CNRS, IRIT

[2]University of California at Berkeley

1. INTRODUCTION

We investigate how to allocate resources when the capacity of the channel fluctuates over time, and when the scheduler does not have full information regarding the actual state of the system. The main motivation comes from communication networks in which the available capacity might evolve over time. A nice introduction to the role of information in scheduling, mostly in the context of stability, is provided in [1]. This paper shows how the problem can be cast as a Partially Observable MDP (POMDP), and observes that in full generality there is very little that can be said. We thus set out to study a simple, yet interesting, instance of the problem in which the channel capacity can only take good values, ON, or bad values OFF. When the channel is ON, data can be transmitted through the channel, but not when it is OFF. Even though the scheduler does not have access to the state of every channel, it does observe successful transmissions.

We consider a system with K channels. Time is discrete, and we let p_k (q_k), $k = 1, \ldots, K$, denote the probability that channel k transitions from ON to OFF (OFF to ON), respectively. This Markov chain model is often referred to as Gilbert-Elliot model in the literature. If a channel is ON and is selected, a successful transmission occurs with probability μ_k. The channel's capacity evolution is positively (negatively) autocorrelated if $1-q_k-p_k > 0$ ($1-q_k-p_k < 0$), respectively. Positive autocorrelation occurs when the probability of remaining in the same state is greater than the probability of transitioning to a different state, whereas negative autocorrelation reflects a tendency to switch states immediately. Negative autocorrelation can naturally arise in real-world systems that operate in discrete time (e.g., slotted communication protocols or periodic scheduling systems) in which states alternate frequently.

We consider *non-preemptive service*, that is, at a decision epoch, only one channel can be selected, and no other channel can be selected until a successful transmission occurs. We denote by $\pi_k(t)$ the belief state, i.e., the probability that channel k is ON at time t, and we let $\mathbb{E}[S_k(\pi_k)]$ denote the expected service time in channel k conditioned on the belief state at the beginning of the transmission being π_k. In particular, $\mathbb{E}[S_k(1-p_k)]$ denotes the expected service time in channel k given a successful transmission occurred in channel k in the previous time slot (the belief state is $1-p_k$ after a successful transmission).

Our main objective is to characterize the scheduling policy that maximizes the long-term number of successful transmissions, i.e., the throughput. With positively autocorrelated channels, the static policy "Serve the Best Channel" (SBC) that serves the channel that maximizes $\mathbb{E}[S_k(1-p_k)]$ is optimal. The negatively autocorrelated case is significantly more difficult. We note that for two symmetric negatively autocorrelated channels, the "Largest Belief State" (LBS) policy that serves at any moment in time the channel $k^* = \arg\max_k(\pi_k(t))$ is optimal.

2. RELATED WORK

The mathematical model [2] is equivalent to ours, except that [2] allows preemption, and we do not. [2] shows that for two symmetric channels and certain conditions on the parameters q, p, and μ, LBS is optimal. For the same (preemptive) model with symmetric channels, [3] provided a closed-form condition under which the LBS policy is optimal for general K. The same authors extended their results in [4] where they considered non-symmetric channels. In a recent paper, Liu et al. [5] consider a similar model to that of [2] and calculate (approximately) the Whittle index. To the best of our knowledge, our work is the first dealing with the non-preemptive setting.

3. MODEL DESCRIPTION

We consider a discrete-time scheduling problem with an infinite number of waiting jobs and K channels, only one of which can be selected, or activated, in each time slot. Channel k has an independent associated Markov chain process X_k describing its environment with two states: 'ON' and 'OFF'. The corresponding transition matrix is $P_k = \begin{bmatrix} 1-p_k & p_k \\ q_k & 1-q_k \end{bmatrix}$. We refer to channel k as positive (negative) autocorrelated when $1-p_k-q_k \geq 0$ ($1-p_k-q_k \leq 0$). The evolution of a channel k is i.i.d. if $1-p_k-q_k = 0$.

The activation process is *non-preemptive*, meaning that once a channel is activated, the job first in line must be served to completion before another activation can occur.

*Corresponding author: thomas.hira@irit.fr

Research partially supported by the French "Agence Nationale de la Recherche (ANR)" through the project ANR-22-CE25-0013-02 (ANR EPLER) and through the France 2030 program Grant NF-NAI: ANR-22-PEFT-0003

Copyright is held by author/owner(s).

As a result, decision times correspond to moments when there is no job in service. An action at decision time t is a vector $u(t) = (u_1(t), ..., u_K(t))$, with $u_k(t) \in \{0, 1\}$ and constraint $\sum_{k=1}^{K} u_k(t) = 1$. We refer to the function u as a policy. Activating a channel with an 'ON' environment causes a service completion (i.e., yields a reward of 1) with probability μ_k, while any other scenario results in no reward.

The controller cannot observe the states of the environments X_k, but it can observe service completions, so we have a Partially Observable Markov Decision Process (POMDP) formulation. The controller's belief state at time t is denoted by $\pi(t) = (\pi_1(t), \ldots, \pi_K(t)) \in [0, 1]^K$, where each $\pi_k(t)$ is the controller's belief probability of channel k being in the 'ON' state. How the belief states are updated will be made precise in Section 4.1. The reward at time t for channel k with belief state π_k and action u_k is defined as

$$R_t\big(\pi_k, u_k\big) := \begin{cases} 1 \text{ with probability } \pi_k \mu_k u_k, \\ 0 \text{ otherwise.} \end{cases}$$

Define U as the space of non-preemptive policies $u(\cdot)$. We denote by g^u the expected average reward (the transmission rate or throughput) of a policy $u \in U$, that is,

$$g^u := \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1}\sum_{k=1}^{K} R_t(\pi_k(t), u_k(t))\right],$$

and by g^* the optimal expected average reward, $g^* := \max_{u\in U} g^u$. Our objective is to find a policy that is in the set of average-optimal policies, $U^* := \{u \in U \mid g^u = g^*\}$.

We define the times $t_0, t_1, \ldots,$ which are random variables depending on the policy u employed:

$$\begin{cases} t_0 = 0, \\ t_{n+1} = \inf\{t > t_n \mid \sum_{k=1}^{K} R_{t-1}(\pi_k(t-1), u_k(t-1)) = 1\}. \end{cases}$$

We refer to these times $t_0, t_1, \ldots,$ as *decision times*. In Lemma 2 we will see that idling is never optimal for our model, hence, each successful transmission time will indeed correspond to a decision time.

4. PRELIMINARY RESULTS

4.1 Probability Updating

The belief state is updated as follows:

$$\pi_k(t+1) = \begin{cases} 1 - p_k & \text{if } R_t(\pi_k(t), u_k(t)) = 1 \\ \bar{T}_k(\pi_k(t)) & \text{if } R_t(\pi_k(t), u_k(t)) = 0 \text{ and } u_k(t) = 1 \\ T_k(\pi_k(t)) & \text{if } u_k(t) = 0, \end{cases}$$

with T_k and $\bar{T}_k$ defined below. To explain the first case, note that if a channel's environment is active at time t and there is a service completion, then we know it was in state ON at time t, and the probability that it is in state ON at time $t+1$ is therefore indeed $\mathbb{P}[X_k(t+1) = ON | X_k(t) = ON] = 1 - p_k$.

For the second case, $T_k(\pi_k(t))$ denotes the updated probability of the environment being 'ON', given that it was not activated, i.e.,

$$\begin{aligned} T_k(\pi_k) &:= \mathbb{P}\left[X_k(t+1) = \text{ON} | \pi_k(t) = \pi_k; u_k(t) = 0\right] \\ &= (1 - p_k)\pi_k + q_k(1 - \pi_k) = \pi_k(1 - p_k - q_k) + q_k. \end{aligned}$$

Similarly, $\bar{T}_k$ denotes the updated probability of the environment being 'ON' when the channel was activated but no service completion (departure) was observed, i.e.

$$\bar{T}_k(\pi_k) := \frac{\pi_k(1 - p_k - q_k - \mu_k(1 - p_k)) + q_k}{1 - \mu_k \pi_k}.$$

We note that the updating functions $T_k(\cdot)$ and $\overline{T}_k(\cdot)$ are increasing (decreasing) as a function of π_k when $1 - p_k - q_k > 0$ ($1 - p_k - q_k < 0$).

4.2 Effective Service Time

We denote by effective service time the (random) time until a service completion conditioned on an initial belief state, that is,

$$S_k(\pi_k) := \inf\{s \geq 1 \ ; \ R_s(\bar{T}_k^{s-1}(\pi_k), 1) = 1\}.$$

The decision times for a given policy u are given by $t_{n+1} = t_n + S^u_{t_n}$, where S^u_t denotes the effective service time under policy u when the job starts being served at time t, so $S^u_t \sim S_{k^u(t)}(\pi_{k^u(t)}(t))$, with $k^u(t) := \arg\max_k u_k(t)$.

5. OPTIMAL SCHEDULING POLICIES

In this section, we discuss and characterize optimal scheduling policies. We first recall that our objective is to find policies that minimize the average reward, where the reward received is 1 each time a job departs. Let

$$\bar{S}^u := \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} S^u_{t_n}$$

be the average effective service time under policy u. In the following lemma, we highlight that the average reward is inversely proportional to the average service time.

Lemma 1. *For $u \in U$, it holds that $g^u = \left(\bar{S}^u\right)^{-1}$. As a consequence, a policy minimizing the expected effective service time is average-optimal, i.e., $\tilde{u} = \arg\min_{u\in U} \bar{S}^u \iff \tilde{u} \in U^*$.*

This Lemma allows us to find an average-optimal policy by solving the Bellman equation for the effective service time.

Corollary 1. *If the index $k^*(\pi)$ minimizes the right-hand side of the following Bellman equation:*

$$\begin{aligned} &\phi(\pi) + \bar{S}^* \\ &= \min_k \Big(\mathbb{E}[S_k(\pi_k)] + \mathbb{E}[\phi(T_1^{S_k(\pi_k)}(\pi_1), ..., T_{k-1}^{S_k(\pi_k)}(\pi_{k-1}), \\ &1 - p_k, T_{k+1}^{S_k(\pi_k)}(\pi_{k+1}), ..., T_K^{S_k(\pi_k)}(\pi_K))]\Big), \end{aligned}$$

then, for all decision times t, the policy

$$u^*(t) = \begin{cases} u_k(t) = 1 \text{ if } k = k^*(\pi(t)) \\ u_k(t) = 0 \text{ otherwise} \end{cases}$$

belongs to U^*.

In the search for optimal policies, we first highlight that we can focus on non-idling policies.

Lemma 2. *An average-optimal policy will be non-idling.*

This lemma allows us to conclude that upon departure of a job, an optimal policy immediately chooses a new channel to serve. The question is thus which channel to serve given the current belief states π_k, for all k. Results on optimal control policies will be formulated in the next section.

6. OPTIMALITY RESULTS

In this section, we derive average-optimal policies for both positive and negative autocorrelated channels. Our results show that the structure of the optimal policy is very different depending on the nature of the autocorrelation.

6.1 Positively Autocorrelated channels

When for all k the channel's environments are positively autocorrelated, that is $1 - p_k - q_k \geq 0$, the belief state for a channel to be ON is decreasing in the number of steps since the last departure. Hence, the corresponding expected service time $\mathbb{E}[S_k(T_k^i(1-p_k)]$ is increasing in i. We recall that at a decision moment, a job just departed from a given channel k, hence it's belief state is $1-p_k$ and expected service time $\mathbb{E}[S_k(T_k^0(1-p_k)] = \mathbb{E}[S_k(1-p_k)]$. In this section we state that the policy that always serves the channel with the smallest value of $\mathbb{E}[S_k(1-p_k)]$, which we call the *Serve the Best Channel* (SBC) policy, is average optimal. In case of a tie, the SBC policy will choose one of the best channels and use that choice consistently thereafter.

Proposition 1. *If $(1-p_k-q_k) > 0$ for all k, then the SBC policy, which always serves channel k^*, where*

$$k^* = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmin}} \mathbb{E}[S_k(1-p_k)] = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmin}} \frac{p_k+q_k}{q_k\mu_k},$$

is average-optimal. In case of a tie, chose one of the minimizing channels arbitrary, and use this choice consistently thereafter. The optimal average reward is given by

$$g^* = g^{SBC} = \mu_{k^*}\frac{q_{k^*}}{p_{k^*}+q_{k^*}}.$$

6.2 Two Symmetric and Negatively Autocorrelated Channels

In this section, we assume we have two symmetric channels ($p_k = p, q_k = q$, and $\mu_k = \mu$, for $k = 1, 2$) that are negatively autocorrelated, $1 - p - q < 0$. We also assume that both channels have been activated at some point in the past. Hence, at each decision time t, $\pi(t) = (1-p, T^i(1-p))$, where we label the channel that just had a service completion as channel 1, without loss of generality, and where i is the time since channel 2 last had a service completion.

Due to the symmetry in the channels, combined with the fact that the expected effective service time is decreasing in the belief state, the greedy policy (with respect to $\mathbb{E}(S_k(\pi_k))$) simplifies to serving the channel that has currently the largest belief state. We refer to this policy as the *Largest Belief State* (LBS) policy. Though the greedy policy might not be optimal in general, below we highlight that in the case of two symmetric and negatively autocorrelated channels it is. Indeed, having symmetry in the channels makes that any decision of which channel to serve will reset the corresponding belief state (and future evolution) in the same manner. Hence, there is no trade-off between short term and long term and an optimal policy can simply be greedy (LBS in this setting).

Proposition 2. *If $K = 2$, $\mu_k = \mu$, $p_k = p$ and $q_k = q$, for $k = 1, 2$, $1 - p - q < 0$, then the LBS policy, which for a given belief state π serves channel k^*, where*

$$k^* = \arg\max_k(\pi_k),$$

is average-optimal. Since $K = 2$, LBS simplifies to continuously switching between the two channels.

The challenge in extending the optimality of LBS beyond $K = 2$ lies in the fact that, after following LBS in a single decision epoch, the system transitions to a "worse" state (in terms of belief states) compared to what could result from following a sub-optimal policy (due to negative autocorrelated channels). Proposition 2, show that for $K = 2$ the gain achieved outweighs the losses incurred, but we were unable to generalize this result to arbitrary $K > 2$.

Remark 1 (Non-optimality of Greedy Policy with Asymmetric Negatively Autocorrelated Channels)**.** *With symmetric and negatively autocorrelated channels, it is stated in Proposition 2 that the greedy policy, which reduces to LBS in this case, was average optimal. It can be shown (omitted for lack of space) that the greedy policy is not necessarily optimal for non-symmetric negative autocorrelated channels.*

7. CONCLUSION

The analysis of this paper shows that the problem of transmitting with partial channel information in a non-preemptive setting is a challenging one. We obtained optimal policies for positively correlated channels with arbitrary K, and for symmetric negatively autocorrelated channels with $K = 2$. As illustrated in Remark 1, the optimality of the greedy policy for the *symmetric* negative autocorrelation, does not carry over to *non-symmetric* negative autocorrelated channels. Characterizing an optimal policy in more general settings might be out of reach and is left for future research.

8. REFERENCES

[1] A. Asanjarani and Y. Nazarathy, "The role of information in system stability with partially observable servers," *Methodology and Computing in Applied Probability*, vol. 22, pp. 949–968, 2020.

[2] K. Liu, Q. Zhao, and B. Krishnamachari, "Dynamic multichannel access with imperfect channel state detection," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2795–2808, 2010.

[3] K. Wang, L. Chen, Q. Liu, and K. Al Agha, "On optimality of myopic sensing policy with imperfect sensing in multi-channel opportunistic access," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3854–3862, 2013.

[4] K. Wang, L. Chen, and Q. Liu, "On optimality of myopic policy for opportunistic access with nonidentical channels and imperfect sensing," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 5, pp. 2478–2483, 2013.

[5] K. Liu, R. Weber, and C. Zhang, "Low-complexity algorithm for restless bandits with imperfect observations," *Mathematical Methods Operations Research*, vol. 100, pp. 467–508, 2024.