



**HAL**  
open science

# Minimizing the Age of Incorrect Information for Real-time Tracking of Markov Remote Sources

Saad Kriouile, Mohamad Assaad

► **To cite this version:**

Saad Kriouile, Mohamad Assaad. Minimizing the Age of Incorrect Information for Real-time Tracking of Markov Remote Sources. 2021 IEEE International Symposium on Information Theory (ISIT), Jul 2021, Melbourne, Australia. 10.1109/ISIT45174.2021.9518209 . hal-03420460

**HAL Id: hal-03420460**

**<https://hal.science/hal-03420460v1>**

Submitted on 9 Nov 2021

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

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

# Minimizing the Age of Incorrect Information for Real-time Tracking of Markov Remote Sources

Saad Kriouile and Mohamad Assaad

Laboratoire des Signaux et Systèmes, CentraleSupélec, Université Paris-Saclay, 91192 Gif sur Yvette, France

**Abstract**—The age of Incorrect Information (AoII) has been introduced to address the shortcomings of the standard Age of information metric (AoI) in real-time monitoring applications. In this paper, we consider the problem of monitoring the states of remote sources that evolve according to a Markovian Process. A central scheduler selects at each time slot which sources should send their updates in such a way to minimize the Mean Age of Incorrect Information (MAoII). The difficulty of the problem lies in the fact that the scheduler cannot know if the information at side of the monitor is correct or not before receiving the updates and it has then to estimate it. We show that the problem can be modeled as a partially Observable Markov Decision Process Problem framework. We develop a new scheduling scheme based on Whittles index policy. The scheduling decision is made by updating a belief value of the states of the sources, which is to the best of our knowledge has not been considered before in the Age of Information area. To that extent, we proceed by using the Lagrangian Relaxation Approach, and prove that the dual problem has an optimal threshold policy. Building on that, we show that the problem is indexable and compute the expressions of the Whittles indices. Finally, we provide some numerical results to highlight the performance of our derived policy compared to the classical AoI metric.

## I. INTRODUCTION

The notable advance in wireless technology and the availability of low-cost hardware have led to the emergence of real-time monitoring services. In these systems, the monitor needs to know the status of one or multiple processes observed by remote sources. Specifically, the sources send packets that contain the information about the process of interest to the monitor to perform a given task. To that extent, the main goal in these applications is to keep the monitor up to date by receiving the fresh information from different sources. This concept of freshness is captured by the Age of Information (AoI) which is introduced for the first time in [1]. Since then, the AoI has become a hot research topic, and a considerable number of research works have been published on the subject [2]–[10]. Although this metric quantifies the information time lag at the monitor, it fails to capture the correctness of the information at the monitor side. Specifically, the evolution of this metric doesn't take into consideration the state of the information at the monitor side. This has been confirmed in [11] where the authors establish that minimizing AoI gives a sub-optimal policy in minimizing the status error in remotely estimating Markovian sources. To deal with this issue, some works propose to minimize the estimation error or the mean square error ([12] and reference therein). However, the metrics developed in these works are unable to capture the concept of

freshness. In other words, there is no penalty incurred to the monitor or the central entity for being in incorrect state for a long time. A new and relevant metric has been proposed by [13] called Cumulative Marginal Error. This metric accumulates the error from the time of the last received update to the current time. However, this metric doesn't match well with the notion of freshness. In fact, having high cumulative marginal error doesn't imply necessarily that the information at side of the monitor is outdated.

To meet the timeliness requirement in the process estimation framework, the authors in [14] have designed a new metric dubbed as Age of incorrect information AoII that captures the freshness of the information while taking into account the information content acquired from the transmitter. This metric is adopted in the context where a given source is represented as a process denoted by  $X(t)$  and the transmitter send status update to the receiver to inform it about the current state of  $X(t)$ . Under energy or transmission rate constraint, the transmitter cannot use at each time slot the channel to transmit the packet. In this case, as long as the transmitter is in the idle mode, the monitor keeps the last information which may be in the erroneous state compared to the current state of the process  $X(t)$ . Denoting by  $\hat{X}(t)$  the estimated state in the monitor side, being at the incorrect state or equivalently  $\hat{X}(t) \neq X(t)$ , is clearly an undesirable situation with regards to the monitor, and therefore a penalty should be paid. The metric developed by [14], AoII matches with this notion of penalty. Specifically, unlike AoI, AoII grows only if the estimated state  $\hat{X}(t)$  in the monitor side is different from the real state of the process of interest  $X(t)$ . While if  $X(t) = \hat{X}(t)$ , unlike Cumulative marginal error, the AoII will be equal to 0. To that extent in [14], [15], the authors consider the problem of minimizing the average AoII in a transmitter-receiver pair scenario where packets are sent over an unreliable channel subject to a transmission rate constraint. They derive the optimal solution which is of the form threshold-based policy. The work in [16] studies the AoII metric in the simple context of monitoring a symmetric binary information source over a delay system with feedback. The authors proposes a dynamic programming algorithm to compute the optimal sampling policy.

However, [14]–[16] assume that the scheduler has a perfect knowledge about the process  $X(t)$  at each time slot  $t$  and restrict the analysis to one transmitter-receiver pair communication. On the opposite of that, in this paper, we tackle a realistic case in which a scheduler tracks the states of multiple

remote sources and selects at each time a subset of them to which the updates are sent, in such a way to minimize the Mean Age of Incorrect Information (MAoII). Furthermore, the scheduler does not know the instantaneous state of the remote sources until it receives their updates. Specifically, our contributions can be summarized as follows:

- Since the scheduler cannot know at each time the current states of the sources before it receives their updates, it cannot know exactly the value of MAoII and has to track/predict its evolution. To that end, we introduce a belief state at the monitor, which can be interpreted as the probability that the state at the monitor side is correct, i.e.  $\hat{X}(t) = X(t)$ . We then describe how this belief state can be derived and can be used in the development of the scheduling policy.
- We then formulate the MAoII-based scheduling problem and show that it belongs to the family of Restless Multi-Armed Bandit (RMAB) problems. The optimal solution of this type of problem is known to be out of reach. To circumvent this difficulty, we develop the low-complex and efficient policy called Whittle's index policy (WIP) using the Lagrangian Relaxation Approach.

## II. SYSTEM MODEL

### A. Network description

We consider in our paper  $N_u$  users that generate and send status updates about the process of interest to a central entity over unreliable channels. Time is considered to be discrete and normalized to the time slot duration. More specifically, each user  $i$  observes an information process of interest  $X_i(t)$  and at the request of the monitor, it samples the process  $X_i(t)$  and send it to the monitor over an unreliable channel. Based on the last received update, the monitor constructs an estimate of the process, denoted by  $\hat{X}_i(t)$ . We suppose that the packet containing the information of interest, if it is successfully transmitted, will be instantaneously delivered to the monitor. In other words, if the monitor allows a user  $i$  to transmit at time  $t$ , it receives the value of  $X_i(t)$  at the same time  $t$  in the case where the packet is successfully transmitted. Therefore, it updates the estimate process as  $\hat{X}_i(t) = X_i(t)$ . In any other case, namely when the user  $i$  is not authorized to transmit or when the packet is unsuccessfully transmitted, the monitor keeps the same value at time slot  $t$ , specifically  $\hat{X}_i(t) = \hat{X}_i(t-1)$ . As for the unreliable channel, we suppose that for user  $i$ , at each time slot  $t$ , the probability of having successful transmission is  $\rho_i$ , and  $1 - \rho_i$  otherwise. Consequently, the channel realizations are independent and identically distributed (i.i.d.) over time slots that we denote  $c_i(t)$ , i.e.  $c_i(t) = 1$  if the packet is successfully transmitted and  $c_i(t) = 0$  otherwise.

The next aspect of our model that we tackle is the nature of the process  $X_i(t)$ . To that extent, for each user  $i$ , the information process of interest  $X_i(t)$  evolves under Markov chain. For that, we define the probability of remaining at the same state in the next time slot as  $p_i$ . Similarly, the probability

of transitioning to another state is  $r_i$ . Denoting by  $N_i$  the number of possible states of  $X_i(t)$ , then the following always holds:

$$p_i + (N_i - 1)r_i = 1 \quad (1)$$

In this paper, we study the case where  $p_i \geq r_i$ .

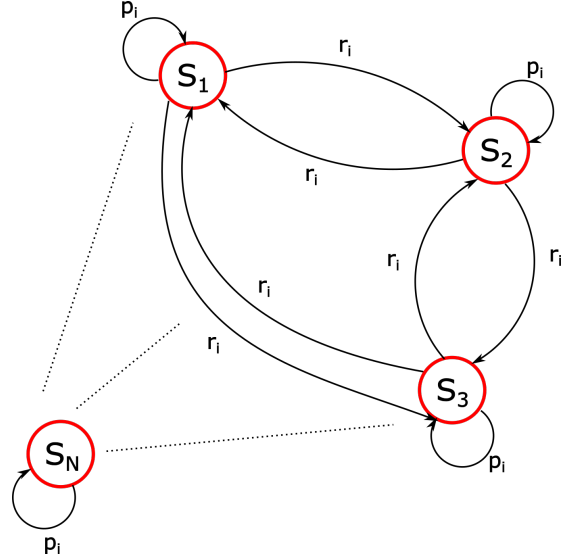


Fig. 1: Illustration of process  $X_i(t)$

### B. Penalty function dynamics

In this paper, we study the mean age of incorrect information (MAoII) penalty function and we compare it with Age of Information metric (AoI). We see how it is relevant, accurate and more realistic to consider MAoII metric compared to AoI metric in order to have a good performance with regard to the empirical value of the age of information. For that purpose, we start by reintroducing in the next section the age of information to emphasize its shortcomings, then we propose as an alternative metric MAoII.

1) *Age of information penalty function*: The standard metric (AoI) that captures the freshness of information for user  $i$  is:

$$\delta_{AoI}(t) = t - g_i(t) \quad (2)$$

where  $g_i(t)$ <sup>1</sup> is the time-stamp of the last successfully received packet by monitor. This metric captures the lifetime of the last update at the monitor without taking into account the correctness of the information. Thereby, this makes it fall short in some applications. For instance in some scenarios, the age will increase but the information of interest remains at the same state. Nevertheless, to further emphasize the shortcoming of this metric, we provide the Whittle index policy considering this metric which is already derived in [14]. And we give some numerical results that show the shortage of this policy.

<sup>1</sup>Considering our system model detailed in II-A,  $g_i(t)$  refers also to the sampling time of the information of interest contained in the last successfully received packet

2) *Mean Age of incorrect information penalty function:*

The age of incorrect information has been introduced the first time in [14]. This metric captures the freshness of informative updates. Specifically, if the monitor acquires the information about the process  $X_i(t)$ , as long as the state of the process  $X_i(t)$  remains at the same state in the next time slots, the age of the incorrect information will not increase, since there is no new information unknown by the monitor. In [14], the authors presume that the scheduler has a perfect knowledge of the process at each time slot and restrict their analysis to a transmitter-receiver pair communication. While in our case, we consider that the monitor which plays the role of the scheduler, knows only the state of the last successively received packet and we extend our analysis to a communication involving several users that can transmit at each time slot. Accordingly, the explicit expression of MAoII metric is:

$$\delta_{MAoII}(t) = \mathbb{E}_{V_i}[(t - V_i(t))] \quad (3)$$

where  $V_i(t)$  is the last time instant such that  $\mathbf{1}_{\{X_i(V_i(t)) = \hat{X}_i(g_i(t))\}} = 1$ .

**Remark 1.** *It is worth mentioning that, as it was explained in Section II-A, the reception of the successfully transmitted packet takes place at time slot  $g_i(t)$ . This means that  $\hat{X}_i(g_i(t)) = X_i(g_i(t))$ .*

In order to use this metric effectively in a partially Observable Markov Decision Process Problem, we need to take into consideration the markovian nature of the process  $X_i(t)$ . To that extent, we introduce in the next section the notion of the belief that represents the probability that  $\hat{X}_i(t)$  is in the correct state.

### C. Metrics evolution

In this section, we describe mathematically the evolution of each metric depending on the system parameters and the action taken. We denote by  $d_i(t)$  the action prescribed to user  $i$  at time slot  $t$  and by  $a_i$ ,  $b_i$ , the age of information and the mean age of incorrect information penalty functions respectively.

1) *AoI:* Considering our system model, the age of information of user  $i$  evolves as follows: If the user  $i$  is scheduled ( $d_i(t+1) = 1$ ), the value of AoI goes to state 0 if the packet is successively transmitted ( $c_i(t+1) = 1$ ), otherwise, the value of AoI is increased by one ( $c_i(t+1) = 0$ ). If the user  $i$  is not scheduled ( $d_i(t+1) = 0$ ), the value of AoI is increased by one. Accordingly, the evolution of the age of the user  $i$  can be summarized in the following:

$$a_i(t+1) = \begin{cases} 0 & \text{if } d_i(t+1) = 1, c_i(t+1) = 1 \\ a_i(t) + 1 & \text{else} \end{cases} \quad (4)$$

As for the second metric, to highlight the notion of correctness, the monitor maintains a belief value  $\pi_i(t)$  which is defined as the probability that the information state in the monitor,

$\hat{X}_i(t) = \hat{X}_i(g_i(t)) = X_i(g_i(t))^2$  at time  $t$  being correct. Explicitly  $\pi_i(t) = Pr(\hat{X}_i(t) = X_i(t))$ . One can show that  $\pi_i(t)$  evolves as follows:

**Lemma 1.**

$$\pi_i(t+1) = \begin{cases} 1 & \text{if } d_i(t+1) = 1, c_i(t+1) = 1 \\ \pi_i(t)p_i + r_i(1 - \pi_i(t)) & \text{else} \end{cases} \quad (5)$$

*Proof:* See appendix A. ■

2) *MAoII:* According to the expression of MAoII given in section II-B2,  $(t - V_i(t))$  is a random variable that we denote  $A_i(t)$  that satisfies:

**Lemma 2.**

$$A_i(t) = \begin{cases} 0 & \text{w.p } \pi_i(t) \\ 1 & \text{w.p } \pi_i(t-1).(1-p_i) \\ 2 & \text{w.p } \pi_i(t-2).(1-p_i).(1-r_i) \\ 3 & \dots \dots \\ \vdots & \\ t - g_i(t) - 1 & \text{w.p } \pi_i(g_i(t) + 1).(1-p_i) \\ & \quad \cdot (1-r_i)^{t-g_i(t)-2} \\ t - g_i(t) & \text{w.p } (1-p_i).(1-r_i)^{t-g_i(t)-1} \end{cases} \quad (6)$$

*Proof:* See appendix B. ■

Therefore, the mean of the age of the incorrect information at slot  $t$  equals to the mean of  $A_i(t)$ , i.e.

$$\begin{aligned} n_i(t) &= \mathbb{E}[A_i(t)] \\ &= \sum_{k=0}^{t-g_i(t)-1} k(1-p_i)(1-r_i)^{k-1}\pi_i(t-k) \\ &\quad + (t-g_i(t)).(1-p_i).(1-r_i)^{t-g_i(t)-1} \\ &= \sum_{k=1}^{t-g_i(t)} (t-g_i(t)-k)(1-p_i)(1-r_i)^{t-g_i(t)-k-1}\pi_i(g_i(t)+k) \\ &\quad + (t-g_i(t)).(1-p_i).(1-r_i)^{t-g_i(t)-1} \end{aligned} \quad (7)$$

One can establish that for all  $t$ , using definition of  $g_i(t)$ ,  $\pi_i(g_i(t)) = 1$ . Hence, according to the evolution of  $\pi_i(\cdot)$  in Lemma 1, for all  $k \leq t - g_i(t)$ ,  $\pi_i(g_i(t) + k)$  depends only on  $k$  and  $i$ . More precisely, we have that for each  $k \leq t - g_i(t)$ ,  $\pi_i(g_i(t) + k) = \pi_i^k$  where  $\pi_i^k$  is a sequence defined by induction as follows:

$$(\pi_i^k) = \begin{cases} \pi_i^0 = 1 \\ \pi_i^{k+1} = p_i\pi_i^k + r_i(1 - \pi_i^k) & \text{if } k \geq 0 \end{cases} \quad (8)$$

In light of that fact, we have that:

$$n_i(t) = \sum_{k=0}^{t-g_i(t)} (t-g_i(t)-k)(1-p_i)(1-r_i)^{t-g_i(t)-k-1}\pi_i^k \quad (9)$$

<sup>2</sup>We recall that as long as the monitor has not received any new update from the source at time instant  $t$ , it maintains the last update successively received at time instant  $g_i(t)$ . In other words,  $\hat{X}_i(t) = \hat{X}_i(g_i(t))$ .

We conclude that  $n_i(t)$  depends on  $t - g_i(t)$  and  $i$ . Therefore, we let  $n_i(t) \triangleq n_i(t - g_i(t))$ .

To that extent, at time slot  $t+1$ , if the user  $i$  is scheduled and the packet is successively transmitted, then  $g_i(t+1) = t+1$ . Accordingly, at time slot  $t+1$ , MAoII equals to  $n_i(t+1 - g_i(t+1)) = n_i(0)$ . If the user  $i$  is not scheduled or if the packet is not successively transmitted, then  $g_i(t+1) = g_i(t)$ . Therefore, MAoII will transit to  $n_i(t+1 - g_i(t+1)) = n_i(t - g_i(t) + 1)$ . Based on this and denoting  $j(t)$  the index such that  $n_i(j(t))$  is the value of MAoII at time slot  $t$ , MAoII will transit to the value  $n_i(j(t) + 1)$  at time instant  $t + 1$ . To sum up, the evolution of MAoII can be summarized as follows:

$$b_i(t+1) = \begin{cases} n_i(0) & \text{if } d_i(t+1) = 1, c_i(t+1) = 1 \\ n_i(j(t) + 1) & \text{else} \end{cases} \quad (10)$$

where  $b_i(t) = n_i(j(t))$ .

### III. PROBLEM FORMULATION

In this section, we consider a given metric denoted by  $m$  where  $m$  can be either AoI or MAoII. We denote further by  $N_u$  the total number of users in the system. We let the vector  $\mathbf{m}$  at time  $t$  be  $\mathbf{m}(t) = (m_1(t), \dots, m_{N_u}(t))$  where  $m_i(t)$  is the penalty function at the central entity of user  $i$  with respect to the metric  $m$  at time slot  $t$ . Our aim is to find a scheduling policy that allocates per each time slot, the available channels ( $M$  channels) to a given subset of users ( $M$  users,  $M \leq N_u$ ) in a such way to minimize the total expected average penalty function of the metric considered. A scheduling policy  $\phi$  is defined as a sequence of actions  $\phi = (\mathbf{d}^\phi(0), \mathbf{d}^\phi(1), \dots)$  where  $\mathbf{d}^\phi(t) = (d_1^\phi(t), d_2^\phi(t), \dots, d_{N_u}^\phi(t))$  is a binary vector such that  $d_i^\phi(t) = 1$  if the user  $i$  is scheduled at time  $t$ . Denoting by  $\Phi$ , the set of all causal scheduling policies, then our scheduling problem can be formulated as follows:

$$\begin{aligned} & \text{minimize}_{\phi \in \Phi} \quad \limsup_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}^{\phi \in \Phi} \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N_u} m_i^\phi(t) | \mathbf{m}(0) \right) \\ & \text{subject to} \quad \sum_{i=1}^{N_u} d_i^\phi(t) \leq \alpha N_u \quad t = 1, 2, \dots \end{aligned} \quad (11)$$

where  $\alpha N_u = M$ . The problem in (11) falls into Multi-armed bandit problems and especially Restless Bandit framework. RMAB problems are known to be generally difficult to solve them as they are PSPACE-Hard [17]. To circumvent this complexity, a well-known heuristic is proposed for these types of problems called Whittle's index policy [18]. This policy is based on a Lagrangian relaxation, and was shown to have remarkable performance in real-life applications. To that extent, we tackle in more depth in the next section the Lagrangian relaxation approach applied to our RBP problem. Then, we provide the theoretical analysis to get the low-complex policy: Whittle index policy that we denote by WIP for the two metrics namely, AoI and MAoII.

## IV. LAGRANGIAN RELAXATION AND WHITTLE'S INDEX

### A. Relaxed problem

The Lagrangian relaxation technique is the key component for defining the Whittle's index scheduling policy. First, it consists of relaxing the constraint on the available resources by letting it be satisfied on average rather than in every time slot. More specifically, we define our Relaxed Problem (RP) as follows:

$$\begin{aligned} & \text{minimize}_{\phi \in \Phi} \quad \limsup_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}^\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N_u} m_i^\phi(t) | \mathbf{m}(0) \right) \\ & \text{subject to} \quad \limsup_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}^\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N_u} d_i^\phi(t) \right) \leq \alpha N_u \end{aligned} \quad (12)$$

The Lagrangian function  $f(W, \phi)$  of the problem (12) is defined as:

$$\limsup_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}^\phi \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N_u} m_i^\phi(t) + W d_i^\phi(t) | \mathbf{m}(0) \right) - W \alpha N_u \quad (13)$$

where  $W \geq 0$  can be seen as a penalty for scheduling users. Thus, by following the Lagrangian approach, our next objective is to solve the following problem:

$$\min_{\phi \in \Phi} f(W, \phi) \quad (14)$$

As the term  $W \alpha N_u$  is independent of  $\phi$ , it can be eliminated from the analysis. Baring that in mind, we present the steps to obtain the Whittle's index policy:

- 1) We focus on the one-dimensional version of the problem in (14). Indeed, it can be shown that the  $N_u$ -dimensional problem can be decomposed into  $N_u$  one-dimensional problems that can be solved independently [19]. Accordingly, we drop the user's index for ease of notation involving all user's parameters, and we deal with the one-dimensional problem:
- $$\min_{\phi \in \Phi} \limsup_{T \rightarrow +\infty} \frac{1}{T} \mathbb{E}^\phi \left( \sum_{t=0}^{T-1} m^\phi(t) + W d^\phi(t) | m(0) \right) \quad (15)$$
- 2) We give the structural results on the optimal solution of the one-dimensional problem.
  - 3) We establish the indexability property, which ensures the existence of the Whittle's indices.
  - 4) We derive a closed-form expression of the Whittle's index and, thereby, define the proposed scheduling policy (WIP) for the original problem (11).

### B. Structural results

The problem in (15) can be viewed as an infinite horizon average cost Markov decision process that is defined as follows:

- **States:** The state of the MDP at time  $t$  is the penalty function  $m(t)$ .
- **Actions:** The action at time  $t$ , denoted by  $d(t)$ , specify if the user is scheduled (value 1) or not (value 0).

- **Transitions probabilities:** The transitions probabilities between the different states.
- **Cost:** We let the instantaneous cost of the MDP,  $C(m(t), d(t))$ , be equal to  $m(t) + Wd(t)$ .

The optimal policy  $\phi^*$  of the one-dimensional problem (15) can be obtained by solving the following Bellman equation for each state  $m$ :

$$\begin{aligned} & \theta + V(m) \\ &= \min_{d \in \{0,1\}} \left\{ m + Wd + \sum_{m' \in A^m} \Pr(m \rightarrow m' | d) V(m') \right\} \end{aligned} \quad (16)$$

where  $\Pr(m \rightarrow m' | d)$  is the transition probability from state  $m$  to  $m'$  under action  $d$ ,  $\theta$  is the optimal value of the problem,  $V(m)$  is the differential cost-to-go function and  $A^m$  is the set of states of the metric  $m$ . There exist several numerical algorithms that are developed to solve (16), such as the value iteration algorithm. This later consists first of updating per each iteration the value function  $V_t(\cdot)$  following the recurrence relation for each state  $m$ :

$$\begin{aligned} & \theta + V_{t+1}(m) \\ &= \min_{d \in \{0,1\}} \left\{ m + Wd + \sum_{m' \in A^m} \Pr(m \rightarrow m' | d) V_t(m') \right\} \end{aligned} \quad (17)$$

Given that  $V_0(\cdot) = 0$ , we compute  $V(\cdot)$  exploiting the fact that  $\lim_{t \rightarrow +\infty} V_t(m) = V(m)$  (see [20, Chapter 8.5]). The main shortcoming of this algorithm that it requires high memory and computational complexity. To overcome this complexity, rather than computing the value of  $V(\cdot)$  for all states, we limit ourselves to study the structure of the optimal scheduling policy by exploiting the fact that  $\lim_{t \rightarrow +\infty} V_t(m) = V(m)$ . In that way, we show that the optimal solution of Problem (16) is a threshold-based policy:

**Definition 1.** A threshold policy is a policy  $\phi \in \Phi$  for which there exists  $n$  such that when the current state  $m < n$ , the prescribed action is  $d^- \in \{0,1\}$ , and when  $m \geq n$ , the prescribed action is  $d^+ \in \{0,1\}$  while baring in mind that  $d^- \neq d^+$ .

To that extent, we show that for both metrics considered in our paper, namely AoI and MAoII, the optimal policy of (16) is a threshold based policy. To that end, we specify first the states space  $A^m$  for both metrics, then we provide the expression of the corresponding Bellman equation (16). After that, we establish our desired result.

1) *AoI*: According to Section II-C1,  $a(t)$  evolves in the state space:

$$A^a = \{a^j : j \geq 0, a^j = j\} \quad (18)$$

The corresponding Bellman equation is:

$$\begin{aligned} \theta_a + V(a^j) &= \min \left\{ a^j + V(a^{j+1}); \right. \\ & \quad \left. a^j + W + \rho V(a^0) + (1 - \rho)V(a^{j+1}) \right\} \end{aligned} \quad (19)$$

The analysis have been already done in [2] regarding this metric. Effectively, in [2], the authors demonstrate that the structure of the optimal policy of Problem (15) is a threshold based policy. They prove further that this policy is increasing with the age. i.e.:

**Proposition 1.** When  $m = a$ , the optimal solution of the problem in (15) is an increasing threshold policy. Explicitly, there exists  $n$  such that when the current state  $a^j < a^n$ , the prescribed action is a passive action, and when  $a^j \geq a^n$ , the prescribed action is an active action.

One can see the detailed proof in [2].

2) *MAoII*: According to Section II-C2,  $b(t)$  evolves in the state space:

$$A^b = \{b^j : j \geq 0, b^j = \sum_{k=0}^j k(1-p)(1-r)^{k-1} \pi^{j-k}\} \quad (20)$$

Therefore, the expression of Bellman equation at state  $b^j$

$$\begin{aligned} \theta_b + V(b^j) &= \min \left\{ b^j + V(b^{j+1}); \right. \\ & \quad \left. b^j + W + \rho V(b^0) + (1 - \rho)V(b^{j+1}) \right\} \end{aligned} \quad (21)$$

**Theorem 1.** When  $m = b$ , the optimal solution of the problem in (15) is an increasing threshold policy. Explicitly, there exists  $n$  such that when the current state  $b^j < b^n$ , the prescribed action is a passive action, and when  $b^j \geq b^n$ , the prescribed action is an active action.

*Proof:* The proof can be found in Appendix C. ■

### C. Indexability and Whittle's index expressions

In order to establish the indexability of the problem and find the Whittle's index expressions, we provide the steady-state form of the problem in (15) under a threshold policy  $n$ . Explicitly:

$$\underset{n \in \mathbb{N}^*}{\text{minimize}} \quad \overline{m}^n + W\overline{d}^n \quad (22)$$

where  $\overline{m}^n$  is the average value of the penalty function with the respect to the metric  $m$ , and  $\overline{d}^n$  is the average active time under threshold policy  $n$ . Specifically:

$$\overline{m}^n = \lim_{T \rightarrow +\infty} \sup \frac{1}{T} \mathbb{E}^n \left( \sum_{t=0}^{T-1} m(t) | m(0), tp(n) \right) \quad (23)$$

$$\overline{d}^n = \lim_{T \rightarrow +\infty} \sup \frac{1}{T} \mathbb{E}^n \left( \sum_{t=0}^{T-1} d(t) | m(0), tp(n) \right) \quad (24)$$

where  $tp(n)$  denotes the threshold policy  $n$ . With the intention of computing  $\overline{m}^n$  and  $\overline{d}^n$ , we derive the stationary distribution of the Discrete Time Markov Chain, DTMC that represents the evolution of MAoII under threshold policy  $n$ . One can show that the steady state distribution in question is the same for both metrics, AoI and MAoII. Specifically:

**Proposition 2.** For  $m = a, b$ , for a given threshold  $n$ , the DTMC admits  $u^n(m^j)$  as its stationary distribution:

$$u^n(m^j) = \begin{cases} \frac{\rho}{n\rho+1} & \text{if } 0 \leq j \leq n \\ (1-\rho)^{j-n} \frac{\rho}{n\rho+1} & \text{if } j \geq n+1 \end{cases} \quad (25)$$

*Proof:* The proof can be found in Appendix D. ■

By exploiting the above results, we can now proceed with finding a closed-form of the average cost of any threshold policy.

**Proposition 3.** For a given threshold  $n$ , the average cost of the threshold policy is  $\bar{m}^n$ :

- $m = a$ :

$$\bar{a}^n = \frac{[(n-1)^2 + (n-1)]\rho^2 + 2\rho(n-1) + 2}{2\rho(n\rho+1)} \quad (26)$$

- $m = b$ :

$$\begin{aligned} \bar{b}^n = & \frac{\rho}{n\rho+1} \left[ \frac{n(N-1)}{Nr} - \frac{(1-Nr)^{n+2}}{(Nr)^2} \right. \\ & + \frac{(1-r)^{n+2}}{r^2} + \frac{(1-\rho)(1-Nr)^{n+2}}{Nr(1-(1-\rho)(1-r))} \\ & \left. - \frac{(1-\rho)(1-r)^{n+2}}{r(1-(1-\rho)(1-r))} + C \right] \quad (27) \end{aligned}$$

where  $C = \frac{(1-Nr)^2}{(Nr)^2} - \frac{(1-r)^2}{r^2} + \frac{(N-1)(1-\rho)}{Nr\rho}$ .

*Proof:* By leveraging the results of Proposition 2 and using the expression of  $m^j$  for  $j \geq 0$ , by definition of  $\bar{m}^n$  given in (23), we get after algebraic manipulations the desired results. ■

**Proposition 4.** For any given threshold  $n$ , the active average time is  $\bar{d}^n$ :

$$\bar{d}^n = \frac{1}{n\rho+1} \quad (28)$$

*Proof:* Likewise, exploiting the results in Proposition 2 and according to the expression (24), we obtain the desired results. ■

To ensure the existence of the Whittle's indices, we need first to establish the indexability property for all users. To that end, we first formalize the indexability and the Whittle's index in the following definitions. We note that in the sequel, we precise the indices of users to differentiate between them.

**Definition 2.** Considering Problem (15) for a given  $W$  and a given user  $i$ , we define  $D_i^m(W)$  as the set of states in which the optimal action (with respect to the optimal solution of Problem (15) considering the metric  $m$ ) is the passive one. In other words,  $m_i^n \in D_i^m(W)$  if and only if the optimal action at state  $m_i^n$  is the passive one.

$D_i^m(W)$  is well defined for both metrics as the optimal solution of Problem (15) is a stationary policy, more precisely, a threshold based policy.

**Definition 3.** A class is indexable if the set of states in which the passive action is the optimal action increases with  $W$ ,

that is,  $W' < W \Rightarrow D_i^m(W') \subseteq D_i^m(W)$ . When the class is indexable, the Whittle's index in state  $m_i^n$  is defined as:

$$W(m_i^n) = \min\{W | m_i^n \in D_i^m(W)\} \quad (29)$$

**Proposition 5.** For each user  $i$ , the one-dimensional problem is indexable for both metrics.

*Proof:* The proof rests on the decrease of  $\bar{d}_i^n$  with  $n$ . One can see [2] for a detailed proof. ■

As the indexability property has been established in the above proposition, we can now assert the existence of the Whittle's index.

**Theorem 2.** For any user  $i$  and state  $m_i^n$ , the Whittle's index is:

- $m = a$

$$W_i(a_i^n) = \frac{n(n+1)\rho_i}{2} + n + 1 \quad (30)$$

- $m = b$

$$\begin{aligned} W_i(b_i^n) = & \frac{(N_i-1)(N_i+1-N_i r_i)\rho_i}{N_i^2 r_i^2} \\ & + (1-N_i r_i)^{n+2} \left( n\rho_i + 1 + \frac{\rho_i}{N_i r_i} \right) \\ & \times \left[ \frac{1-(1-\rho_i)(1+(N_i-1)r_i)}{N_i r_i (1-(1-\rho_i)(1-r_i))} \right] \\ & - (1-r_i)^{n+2} \left( n\rho_i + 1 + \frac{\rho_i}{r_i} \right) \\ & \times \left[ \frac{\rho_i}{r_i (1-(1-\rho_i)(1-r_i))} \right] \quad (31) \end{aligned}$$

*Proof:* The proof can be found in Appendix E. ■

Based on the above proposition, we provide in the following the Whittle's index scheduling policy for the original problem (11).

---

**Algorithm 1** Whittle's index scheduling policy

---

- 1: At each time slot  $t$ , compute the Whittle's index of all users in the system using the expressions given in Proposition 2.
  - 2: Allocating the  $M$  channels to the  $M$  users having the highest Whittle's index values at time  $t+1$ .
- 

## V. NUMERICAL RESULTS

Our goal in this section is to compare the average empirical age of incorrect information under the developed Whittle index policy WIP-MAoII to the one under the baseline policy, denoted by WIP-AoI, that considers the standard AoI metric. More precisely, we plot  $C^{\phi, N_u} = \frac{1}{N_u} \lim_{T \rightarrow +\infty} \sup \frac{1}{T} \mathbb{E}^{\phi} \left( \sum_{t=0}^{T-1} \sum_{i=1}^{N_u} m_i^{emp, \phi}(t) | \mathbf{m}^{emp}(0), \phi \right)$  for  $\phi$  equals to WIP-MAoII and WIP-AoI, in function of  $N_u$ , where  $m_i^{emp, \phi}(\cdot)$  evolves as follows:

- If  $m_i^{emp, \phi}(t) = 0$ , then  $\hat{X}_i(t) = X_i(t)$ .
  - If  $\phi_i(t+1) = 1$ : If the packet is successfully transmitted, then  $\hat{X}_i(t+1) = X_i(t+1)$ . In this

case, the real age of information will stay at the state 0. If the packet is unsuccessfully transmitted, then the monitor maintains the last estimated value of the process of interest, i.e.  $\hat{X}_i(t+1) = \hat{X}_i(t)$ . Knowing that  $\hat{X}_i(t) = X_i(t)$ , the probability that  $\hat{X}_i(t+1) = X_i(t+1)$  ( $m_i^{emp,\phi}(t+1) = 0$ ) is  $p_i$  and the probability that  $\hat{X}_i(t+1) \neq X_i(t)$  ( $m_i^{emp,\phi}(t+1) = 1$ ) is  $1 - p_i$ . Therefore, the value of  $m_i^{emp,\phi}(t+1)$  is as follows:

$$m_i^{emp,\phi}(t+1) = \begin{cases} 0 & w.p \quad \rho_i + (1 - \rho_i)p_i \\ 1 & w.p \quad (1 - p_i)(1 - \rho_i) \end{cases} \quad (32)$$

- If  $\phi_i(t+1) = 0$ : The monitor maintains the last estimated value of the process of interest, i.e.  $\hat{X}_i(t+1) = \hat{X}_i(t)$ . Knowing that  $\hat{X}_i(t) = X_i(t)$ , the probability that  $\hat{X}_i(t+1) = X_i(t+1)$  ( $m_i^{emp,\phi}(t+1) = 0$ ) is  $p_i$  and the probability that  $\hat{X}_i(t+1) \neq X_i(t)$  ( $m_i^{emp,\phi}(t+1) = 1$ ) is  $1 - p_i$ . Therefore, the value of  $m_i^{emp,\phi}(t+1)$  is as follows:

$$m_i^{emp,\phi}(t+1) = \begin{cases} 0 & w.p \quad p_i \\ 1 & w.p \quad (1 - p_i) \end{cases} \quad (33)$$

- If  $m_i^{emp,\phi}(t) \neq 0$ , then  $\hat{X}_i(t) \neq X_i(t)$ .
  - If  $\phi_i(t+1) = 1$ : If the packet is successfully transmitted, then  $\hat{X}_i(t+1) = X_i(t+1)$ . Hence the real age of information will move to the state 0. If the packet is unsuccessfully transmitted, then the monitor maintains the last estimated value of the process of interest, i.e.  $\hat{X}_i(t+1) = \hat{X}_i(t)$ . Knowing that  $\hat{X}_i(t) \neq X_i(t)$ , the probability that  $\hat{X}_i(t+1) = X_i(t+1)$  ( $m_i^{emp,\phi}(t+1) = 0$ ) is  $r_i$  and the probability that  $\hat{X}_i(t+1) \neq X_i(t)$  ( $m_i^{emp,\phi}(t+1) = m_i^{emp,\phi}(t) + 1$ ) is  $1 - r_i$ . Therefore, the value of  $m_i^{emp,\phi}(t+1)$  is as follows:

$$m_i^{emp,\phi}(t+1) = \begin{cases} 0 & w.p \quad \rho_i + (1 - \rho) \\ m_i^{emp,\phi}(t) + 1 & w.p \quad (1 - \rho_i) \times (1 - r_i) \end{cases} \quad (34)$$

- If  $\phi_i(t) = 0$ : The monitor maintains the last estimated value of the process of interest, i.e.  $\hat{X}_i(t+1) = \hat{X}_i(t)$ . Knowing that  $\hat{X}_i(t) \neq X_i(t)$ , the probability that  $\hat{X}_i(t+1) = X_i(t+1)$  ( $m_i^{emp,\phi}(t+1) = 0$ ) is  $r_i$  and the probability that  $\hat{X}_i(t+1) \neq X_i(t)$  ( $m_i^{emp,\phi}(t+1) = m_i^{emp,\phi}(t) + 1$ ) is  $1 - r_i$ . Therefore, the value of  $m_i^{emp,\phi}(t+1)$  is as follows:

$$m_i^{emp,\phi}(t+1) = \begin{cases} 0 & w.p \quad r_i \\ m_i^{emp,\phi}(t) + 1 & w.p \quad (1 - r_i) \end{cases} \quad (35)$$

We consider two scenarios of the network settings:

- 1) For the first scenario, we consider two classes with the respective parameters:

- Class 1:  $\rho_1 = 0.7$ ,  $N_1 = 8$ ,  $r_1 = 0.1^3$ .

<sup>3</sup>The value of  $p_i$  can be directly deduced from Equation 1

- Class 2:  $\rho_2 = 0.5$ ,  $N_2 = 2$ ,  $r_2 = 0.4$ .

- 2) Regarding the second scenario, to shed light on the importance of taking into account the source parameters namely,  $p_i$ ,  $r_i$  and  $N_i$ , in the derivation of Whittle's indices, we consider that the two classes share the same channel statistics, specifically  $\rho_1 = \rho_2$ , while they don't have the same source parameters. To that extent, we consider the following case:

- Class 1:  $\rho_1 = 0.2$ ,  $N_1 = 10$ ,  $r_1 = 0.05$
- Class 2:  $\rho_2 = 0.2$ ,  $N_2 = 3$ ,  $r_2 = 0.3$

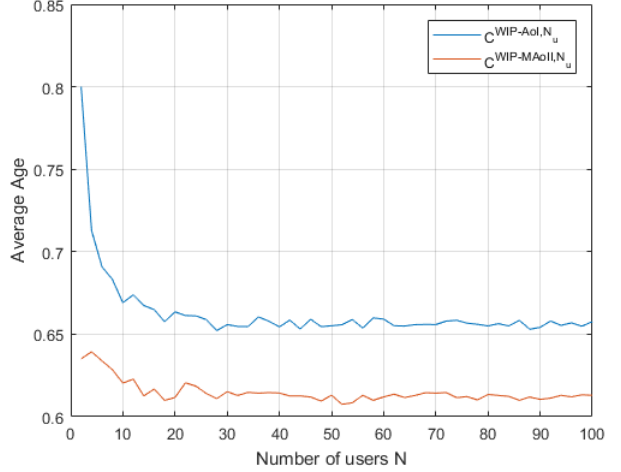


Fig. 2: Comparison between WIP-MAoII and WIP-AoI in terms of the empirical average age: different channel statistics

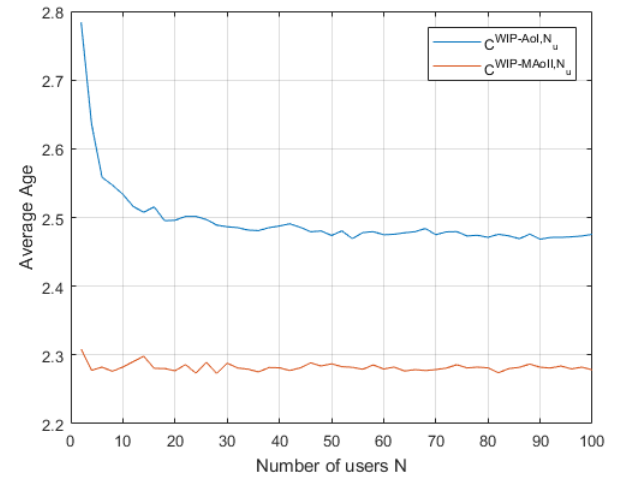


Fig. 3: Comparison between WIP-MAoII and WIP-AoI in terms of the empirical average age: same channel statistics

One can observe that effectively WIP-MAoII gives us better performance than WIP-AoI in terms of minimizing the average empirical age of incorrect information. As consequence, our derivation of Whittle's indices in the Markovian source



framework turns out to be relevant in terms of tracking the real state of remote sources.

## VI. CONCLUSION

In this paper, we considered the problem of remote monitoring of multiple sources where a central entity selects at each time a subset of the sources to send their updates, in such a way to minimize the MAoII metrics. Since the scheduler is unaware of the current state of the source, we introduce a belief state at the monitor in order to predict the evolution of the states of the sources and to derive an estimation of the MAoII. We then developed an efficient scheduling policy based on Whittle's index framework. Finally, we have provided numerical results that highlight the performance of our policy.

## REFERENCES

- [1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 2731–2735.
- [2] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "On the optimality of the whittles index policy for minimizing the age of information," *IEEE Transactions on Wireless Communications*, 2020.
- [3] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Transactions on Mobile Computing*, 2019.
- [4] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2637–2650, 2018.
- [5] P. Zou, O. Ozel, and S. Subramaniam, "Waiting before serving: A companion to packet management in status update systems," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3864–3877, 2019.
- [6] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Optimizing data freshness, throughput, and delay in multi-server information-update systems," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 2569–2573.
- [7] —, "Age-optimal information updates in multihop networks," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 576–580.
- [8] —, "The age of information in multihop networks," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1248–1257, 2019.
- [9] A. Maatouk, Y. Sun, A. Ephremides, and M. Assaad, "Status updates with priorities: Lexicographic optimality," in *2020 18th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2020, pp. 1–8.
- [10] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, "Age-optimal updates of multiple information flows," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*. IEEE, 2018, pp. 136–141.
- [11] Z. Jiang, S. Zhou, Z. Niu, and C. Yu, "A unified sampling and scheduling approach for status update in multiaccess wireless networks," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 208–216.
- [12] Y. Sun, Y. Polyanskiy, and E. Uysal, "Sampling of the wiener process for remote estimation over a channel with random delay," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 1118–1135, 2019.
- [13] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an effective age of information: Remote estimation of a markov source," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*. IEEE, 2018, pp. 367–372.
- [14] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "The age of incorrect information: A new performance metric for status updates," *IEEE/ACM Transactions on Networking*, vol. 28, no. 5, pp. 2215–2228, 2020.
- [15] A. Maatouk, M. Assaad, and A. Ephremides, "The age of incorrect information: an enabler of semantics-empowered communication," *arXiv preprint arXiv:2012.13214*, 2020.

- [16] C. Kam, S. Kompella, and A. Ephremides, "Age of incorrect information for remote estimation of a binary markov source," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*. IEEE, 2020, pp. 1–6.
- [17] 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.
- [18] 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.
- [19] S. Kriouile, M. Larranaga, and M. Assaad, "Asymptotically optimal delay-aware scheduling in wireless networks, arxiv e-prints, p," *arXiv preprint arXiv:1807.00352*, 2018.
- [20] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

## APPENDIX A PROOF OF LEMMA 1

We start by proving the first statement: if  $d_i(t+1) = 1$  and  $c_i(t+1) = 1$ , then the central entity acquires the timeless information about the process  $X_i(t+1)$ . Accordingly, it knows the effective state of the process that we denote by  $S$  at time  $t+1$ . Specifically,  $\hat{X}_i(t+1) = X_i(t+1) = S$ . Therefore, the probability that  $X_i(t+1) = \hat{X}_i(t+1) = S$  knowing the fact that  $X_i(t+1) = S$  is one. Consequently,  $\pi_i(t+1) = 1$ . If  $d_i(t+1) = 1$  and  $c_i(t+1) = 0$ , or  $d_i(t+1) = 0$ , then the information state in the monitor side does not change ( $\hat{X}_i(t+1) = \hat{X}_i(t)$ ). Thus, the probability of the event  $X_i(t+1) = \hat{X}_i(t+1)$ :

$$\begin{aligned}
Pr(X_i(t+1) = \hat{X}_i(t+1)) &= Pr(X_i(t+1) = \hat{X}_i(t+1) | X_i(t) = \hat{X}_i(t)) \\
&\quad \times Pr(X_i(t) = \hat{X}_i(t)) \\
&\quad + Pr(X_i(t+1) = \hat{X}_i(t+1) | X_i(t) \neq \hat{X}_i(t)) \\
&\quad \times Pr(X_i(t) \neq \hat{X}_i(t)) \\
&= Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) = \hat{X}_i(t)) \\
&\quad \times Pr(X_i(t) = \hat{X}_i(t)) \\
&\quad + Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t)) \\
&\quad \times Pr(X_i(t) \neq \hat{X}_i(t)) \\
&= Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) = \hat{X}_i(t))\pi_i(t) \\
&\quad + Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t))(1 - \pi_i(t))
\end{aligned} \tag{36}$$

$Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) = \hat{X}_i(t))$  is the probability of remaining at the same state in the next time slot that equals to  $p_i$ . Decomposing  $Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t))$ :

$$\begin{aligned}
Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t)) &= \sum_{S \neq \hat{X}_i(t)} Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t), X_i(t) = S) \\
&\quad \times Pr(X_i(t) = S | X_i(t) \neq \hat{X}_i(t))
\end{aligned} \tag{37}$$

$Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t), X_i(t) = S)$  is the probability of transitioning from  $S$  to  $\hat{X}_i(t) \neq S$  that equals

to  $r_i$ . Hence:

$$\begin{aligned} Pr(X_i(t+1) = \hat{X}_i(t) | X_i(t) \neq \hat{X}_i(t)) \\ = r_i \sum_{S \neq \hat{X}_i(t)} Pr(X_i(t) = S | X_i(t) \neq \hat{X}_i(t)) \\ = r_i \end{aligned} \quad (38)$$

That is, combining the two results:

$$Pr(X_i(t+1) = \hat{X}_i(t+1)) = \pi_i(t)p_i + (1 - \pi_i(t))r_i \quad (39)$$

Therefore, the proof is complete.

#### APPENDIX B PROOF OF LEMMA 2

We have  $A_i(t) = (t - V_i(t))$ . As  $g_i(t)$  is known by the monitor, then it is a fixed constant. Whereas  $V_i(t)$  which represents the last time instant such that  $\mathbf{1}_{\{X_i(V_i(t)) = \hat{X}_i(g_i(t))\}} = 1$  is unknown by the monitor. Accordingly, it is viewed as a random variable by the monitor. By definition of  $g_i(t)$ ,  $\mathbf{1}_{\{X_i(g_i(t)) = \hat{X}_i(g_i(t))\}} = 1$ , then  $V_i(t)$  takes values in  $[g_i(t), t]$ . To that extent, we distinguish between two cases:

1) The event  $\{V_i(t) = g_i(t)\}$  implies that:

- $X_i(g_i(t)) = \hat{X}_i(g_i(t))$
- For all  $j \in [g_i(t) + 1, t]$ ,  $X_i(j) \neq \hat{X}_i(g_i(t))$

By definition of  $g_i(t)$ , the probability of  $\{X_i(g_i(t)) = \hat{X}_i(g_i(t))\}$  is 1.

The probability of  $\{\forall j \in [g_i(t) + 1, t], X_i(j) \neq \hat{X}_i(g_i(t)) | X_i(g_i(t)) = \hat{X}_i(g_i(t))\}$  is  $(1 - r_i)^{t - g_i(t) - 1} (1 - p_i)$ .

Accordingly, the probability of the event  $\{V_i(t) = g_i(t)\}$  is  $(1 - p_i)(1 - r_i)^{t - g_i(t) - 1}$ .

2) For  $k \in [g_i(t) + 1, t]$ , the event  $\{V_i(t) = k\}$  implies that:

- At time  $k$ ,  $X_i(k) = \hat{X}_i(g_i(t))$ .
- For all  $j \in [k + 1, t]$ ,  $X_i(j) \neq \hat{X}_i(g_i(t))$ .

The probability of  $\{X_i(k) = \hat{X}_i(g_i(t)) = \hat{X}_i(k)\}$  is  $\pi_i(k)$ .

The probability of  $\{\forall j \in [k + 1, t], X_i(j) \neq \hat{X}_i(g_i(t)) | X_i(k) = \hat{X}_i(g_i(t))\}$  is  $(1 - r_i)^{t - k - 1} (1 - p_i)$ .

Accordingly, the probability of the event  $\{V_i(t) = k\}$  is  $\pi_i(k)(1 - p_i)(1 - r_i)^{t - k - 1}$ .

Given that  $A_i(t) = t - V_i(t)$ , then  $A_i(t) = k$  implies  $V_i(t) = t - k$ . That is, the probability of the event  $\{A_i(t) = k\}$  is:

- If  $k = t - g_i(t)$ :

$$(1 - r_i)^{t - g_i(t) - 1} (1 - p_i) \quad (40)$$

- If  $k \in [0, t - g_i(t)]$ :

$$\pi_i(t - k)(1 - r_i)^{k - 1} (1 - p_i) \quad (41)$$

This concludes the proof.

#### APPENDIX C PROOF OF THEOREM 1

We provide first an useful lemma.

**Lemma 3.**  $b^j$  is increasing with  $j$

*Proof:* The explicit expression of  $b^j$  is:

$$\begin{aligned} b^j = \frac{N - 1}{1 + r - p} [1 - (j + 1)(1 - r)^j + j(1 - r)^{j+1}] \\ + \frac{1}{1 + r - p} [(p - r)^{j+1} - (j + 1)(1 - r)^j(p - r) \\ + j(1 - r)^{j+1}] \end{aligned} \quad (42)$$

Therefore, after some computations and mathematical analysis, we obtain:

$$b^{j+1} - b^j = Nr[(1 - r)^{j+1} - (p - r)^{j+1}] \quad (43)$$

Given that  $0 \leq p - r \leq 1 - r$ , then  $(p - r)^{j+1} \leq (1 - r)^{j+1}$ . Therefore,  $(1 - r)^{j+1} - (p - r)^{j+1} \geq 0$ . Hence,  $b^j$  is increasing with  $j$ . ■

Based on this lemma, we prove the following lemma.

**Lemma 4.**  $V(\cdot)$  is increasing with  $b^j$ .

*Proof:* We prove the present lemma by induction using the Value iteration equation (17). In fact, we show that  $V_t(\cdot)$  is increasing and we conclude for  $V(\cdot)$ .

As  $V_0(\cdot) = 0$ , then the property holds for  $t = 0$ . If  $V_t(\cdot)$  is increasing with  $b$ , we show that for  $b^j \leq b^i$ ,  $V_{t+1}^0(b^j) \leq V_{t+1}^0(b^i)$  and  $V_{t+1}^1(b^j) \leq V_{t+1}^1(b^i)$  where for each  $k \in \mathbb{N}$ :

$$V_{t+1}^0(b^k) = b^k + V_t(b^{k+1}) \quad (44)$$

$$V_{t+1}^1(b^k) = b^k + W + \rho V_t(b^0) + (1 - \rho)V_t(b^{k+1}) \quad (45)$$

We have that:

$$V_{t+1}^0(b^j) - V_{t+1}^0(b^i) = b^j - b^i + (V_t(b^{j+1}) - V_t(b^{i+1})) \quad (46)$$

According to Lemma 3, given that  $b^j \leq b^i$ , then  $j \leq i$ . That means  $b^{j+1} \leq b^{i+1}$ . Therefore, since  $V_t(\cdot)$  is increasing with  $b^j$ , we have that:

$$V_{t+1}^0(b^j) - V_{t+1}^0(b^i) \leq 0$$

As consequence,  $V_{t+1}^0(\cdot)$  is increasing with  $b^j$ .

In the same way, we have:

$$V_{t+1}^1(b^j) - V_{t+1}^1(b^i) = b^j - b^i + (1 - \rho)(V_t(b^{j+1}) - V_t(b^{i+1}))$$

Hence:

$$V_{t+1}^1(b^j) - V_{t+1}^1(b^i) \leq 0 \quad (47)$$

As consequence,  $V_{t+1}^1(\cdot)$  is increasing with  $b^j$ .

Since  $V_{t+1}(\cdot) = \min\{V_{t+1}^0(\cdot), V_{t+1}^1(\cdot)\}$ , then  $V_{t+1}(\cdot)$  is increasing with  $b^j$ . Accordingly, we demonstrate by induction that  $V_t(\cdot)$  is increasing for all  $t$ . Knowing that  $\lim_{t \rightarrow +\infty} V_t(b^j) = V(b^j)$ ,  $V(\cdot)$  must be also increasing with  $b^j$ . ■

We define:

$$\Delta V(b^j) = V^1(b^j) - V^0(b^j) \quad (48)$$

where  $\lim_{t \rightarrow +\infty} V_t^0(b^j) = V^0(b^j)$  and  $\lim_{t \rightarrow +\infty} V_t^1(b^j) = V^1(b^j)$ . Subsequently,  $\Delta V(b^j)$  equals to:

$$\Delta V(b^j) = \rho \left[ \frac{W}{\rho} + V(b^0) - V(b^{j+1}) \right] \quad (49)$$

According to Lemma 4,  $V(\cdot)$  is increasing with  $b^{j+1}$ . Therefore,  $\Delta V(b^j)$  is decreasing with  $b^j$ . Hence, there exists  $b^n$  such that for all  $b^j \leq b^n$ ,  $\Delta V(b^j) \geq 0$ , and for all  $b^j > b^n$ ,  $\Delta V(b^j) < 0$ . Given that the optimal action for state  $b^j$  is the one that minimizes  $\min\{V^0(\cdot), V^1(\cdot)\}$ , then for all  $b^j \leq b^n$ , the optimal decision is to stay idle since  $\min\{V^0(b^j), V^1(b^j)\} = V^0(b^j)$ , and for all  $b^j > b^n$ , the optimal decision is to transmit since  $\min\{V^0(b^j), V^1(b^j)\} = V^1(b^j)$ . Specifically, as  $b^j$  is increasing with  $j$ , there exists  $n$  such that for all  $j < n$ , the optimal action is passive action, and for all  $j \geq n$ , the optimal action is the active one.

#### APPENDIX D PROOF OF PROPOSITION 2

In order to demonstrate this proposition, we need to resolve the full balance equation under threshold policy  $n$  at each state  $m^j$  for  $m = a$  and  $m = b$ :

$$u^n(m^j) = \sum_{i=0}^{+\infty} pt^n(i \rightarrow j) u^n(m^i) \quad (50)$$

where  $pt^n(i \rightarrow j)$  denotes the transition probability from the state  $m^i$  to state  $m^j$  under threshold policy  $n$ . After some computations, we obtain the desired result which is valid for both  $m = a$  and  $m = b$ .

#### APPENDIX E PROOF OF THEOREM 2

For  $m = a$ , since the analysis are same as in [2], we skip this case for sake of space. When  $m = b$ , using Definition 3 to find the Whittles index expressions can be tricky and difficult. To circumvent this, we first define the sequence  $W_i(b_i^n)$  as the intersection points between  $\overline{b_i^n} + W\overline{d_i^n}$  and  $\overline{b_i^{n+1}} + W\overline{d_i^{n+1}}$ . Explicitly:

$$W_i(b_i^n) = \frac{\overline{b_i^{n+1}} - \overline{b_i^n}}{\overline{d_i^n} - \overline{d_i^{n+1}}} \quad (51)$$

According to the results in [32, Corollary 2.1], if  $W_i(b_i^n)$  is increasing with  $b_i^n$ , then the Whittles index for any state  $b_i^n$  is nothing but  $W_i(b_i^n)$ . To that extent, we prove that  $W_i(b_i^n)$  is increasing with  $b_i^n$ . However, since  $b_i^n$  is increasing with  $n$  when  $m = b$  (Lemma 3), it is sufficient to show that  $W_i(b_i^n)$  is increasing with  $n$  to establish the desired result.

Therefore, we first seek a closed-form expression of the intersection point  $W_i(b_i^n)$ , we obtain:

$$\begin{aligned} W_i(b_i^n) &= \frac{(N_i - 1)(N_i + 1 - N_i r_i) \rho_i}{N_i^2 r_i^2} \\ &\quad + (1 - N_i r_i)^{n+2} \left( n \rho_i + 1 + \frac{\rho_i}{N_i r_i} \right) \\ &\quad \times \left[ \frac{1 - (1 - \rho_i)(1 + (N_i - 1)r_i)}{N_i r_i (1 - (1 - \rho_i)(1 - r_i))} \right] \\ &\quad - (1 - r_i)^{n+2} \left( n \rho_i + 1 + \frac{\rho_i}{r_i} \right) \\ &\quad \times \left[ \frac{\rho_i}{r_i (1 - (1 - \rho_i)(1 - r_i))} \right] \end{aligned} \quad (52)$$

Now, we provide the main result that allows us to affirm that  $W_i(b_i^n)$  is effectively the Whittle index of state  $b_i^n$ .

**Lemma 5.** *The sequence  $W_i(b_i^n)$  is increasing with  $n$ .*

*Proof:* After some mathematical analysis and algebraic manipulations, we get:

$$\begin{aligned} W_i(b_i^{n+1}) - W_i(b_i^n) &= ((n+1)\rho_i + 1)[(1 - r_i)^{n+2} - (1 - N_i r_i)^{n+2}] \\ &\quad + \frac{((n+1)\rho_i + 1)(1 - \rho_i)}{1 - (1 - \rho_i)(1 - r_i)} r_i [N_i(1 - N_i r_i)^{n+2} - (1 - r_i)^{n+2}] \end{aligned} \quad (53)$$

We have that:

$$N_i(1 - N_i r_i)^{n+2} - (1 - r_i)^{n+2} \geq (1 - N_i r_i)^{n+2} - (1 - r_i)^{n+2} \quad (54)$$

Thus:

$$\begin{aligned} W_i(b_i^{n+1}) - W_i(b_i^n) &\geq ((n+1)\rho_i + 1) \times \left[ 1 - \frac{(1 - \rho_i)}{1 - (1 - \rho_i)(1 - r_i)} r_i \right] \\ &\quad \times [(1 - r_i)^{n+2} - (1 - N_i r_i)^{n+2}] \end{aligned} \quad (55)$$

Given that  $(1 - r_i)^{n+2} - (1 - N_i r_i)^{n+2} \geq 0$  and  $1 - \frac{(1 - \rho_i)}{1 - (1 - \rho_i)(1 - r_i)} r_i = \frac{\rho_i}{1 - (1 - \rho_i)(1 - r_i)} \geq 0$ , therefore:

$$W_i(b_i^{n+1}) - W_i(b_i^n) \geq 0 \quad (56)$$

That concludes the proof. ■