An Easier-to-Verify Sufficient Condition for Whittle Indexability and Application to AoI Minimization

Sixiang Zhou and Xiaojun Lin
Elmore Family School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN, USA
{zhou1087@purdue.edu, linx@ecn.purdue.edu }

***Abstract*—We study a scheduling problem for a base-station transmitting status information to multiple user-equipments (UE) with the goal of minimizing the total expected Age-of-Information (AoI). Such a problem can be formulated as a Restless Multi-Armed Bandit (RMAB) problem and solved asymptotically-optimally by a low-complexity Whittle index policy, if each UE's sub-problem is Whittle indexable. However, proving Whittle indexability can be highly non-trivial, especially when the value function cannot be derived in closed-form. In particular, this is the case for the AoI minimization problem with stochastic arrivals and unreliable channels, whose Whittle indexability remains an open problem. To overcome this difficulty, we develop a sufficient condition for Whittle indexability based on the notion of active time (AT). Even though the AT condition shares considerable similarity to the Partial Conservation Law (PCL) condition, it is much easier to understand and verify. We then apply our AT condition to the stochastic-arrival unreliable-channel AoI minimization problem and, for the first time in the literature, prove its Whittle indexability. Our proof uses a novel coupling approach to verify the AT condition, which may also be of independent interest to other large-scale RMAB problems.**

I. INTRODUCTION

Many emerging wireless applications, e.g., Internet-of-Things (IoT) [1] and Intelligent Transportation Systems [2], require fresh information updates to be delivered in real-time. AoI (Age-of-Information), which is defined as the elapsed time of the latest-received information packet since it was generated at the source, can be used to capture the freshness of information. Thus, there have been significant interests in how to schedule wireless transmissions to minimize AoI [3]–[12].

When information updates for multiple users are transmitted over an unreliable wireless channel, the resulting AoI minimization problem can be modelled as a Markov Decision Process (MDP), or more specifically a Restless Multi-Arm Bandit (RMAB) problem [5]–[14]. Unfortunately, such a problem is known to suffer from the curse-of-dimensionality, i.e., the complexity of the problem grows exponentially with the number of users. Recently, Whittle index has been applied to decompose such a problem into multiple per-agent MDP. When each agent's sub-problem is Whittle indexable, the corresponding Whittle index policy is known to be asymptotically optimal when the system size is large [15], [16].

However, establishing Whittle indexability is by itself quite an involved problem. In some cases, the value function of the per-agent problem can be solved in closed-form, based on which one can verify Whittle indexability and even derive a closed-form expression of the Whittle index. This is the case, e.g., when a "generate-at-will" source transmits information over a Bernoulli channel [5]. Part of the reason that the value function can be solved in closed form for this "generate-at-will" setting is that the state space of the per-agent problem is one dimensional (i.e., the state includes only the age). Unfortunately, for slightly more complex settings, the value function can no longer be derived in closed-form. For example, when the information arrivals are stochastic, i.e., new update packets are generated randomly, the state space will be two-dimensional (i.e., the state includes both the age and the latest packet-generation time). It then becomes difficult to derive the value function in closed-form. As pointed out in [11], Whittle indexability under the stochastic-arrival and unreliable-channel settings remains an open problem, even though it seems quite intuitive to be true. (We note that the recent work in [12] claims to provide a closed-form expression for the value function for this stochastic-arrival setting, based on which the authors claim Whittle indexability. However, our own numerical results based on the value iteration indicate that the value-function expressions in [12] are imprecise.) Thus, there is a need to develop new approaches for verifying Whittle indexability that do not require closed-form expressions of the value function, but instead are directly based on the problem structure.

It turns out that in the literature of RMAB and Whittle index [17]–[20], there have been some sufficient conditions for Whittle indexability, most notably the General Conservation Law (GCL) condition in [17] and the Partial Conservation Law (PCL) condition in [18]. These powerful conditions can be checked without knowing the value function. Between them, PCL is more general than GCL, and therefore can potentially be applied to a larger set of problems. However, the application of PCL is scarce in the literature. In particular, to the best of our knowledge, there has been no application of PCL to the AoI minimization problem with stochastic arrivals and unreliable channels. When we study PCL, we found two possible reasons for this lack of application. First, verifying

This work has been partially supported by NSF grants CNS-2113893 and CNS-2225950. Part of this work was completed when the second author moved to The Chinese University of Hong Kong.

the PCL condition is also non-trivial (even for simple toy examples shown in [18]). Second, both the statement of the PCL condition by itself and its proof of sufficiency are not easy to understand, and hence it is unclear how to develop effective approaches to verify the PCL condition.

In this paper, our first contribution is to develop a sufficient condition based on the notion of active time (AT). While the AT condition shares considerable similarity to the PCL condition, there are also a number of crucial differences. For example, there is no need to check whether the per-stage cost is admissible by an adaptive greedy algorithm as in [18]. (See further discussions of their differences in Section 3.) As a result, the AT condition is both easier to understand (and its proof of sufficiency is more straight-forward) and easier to verify. As an example of the power of the AT condition, we show how easily it can be applied to AoI minimization under the generate-at-will setting to establish Whittle indexability, without the need of finding the value function.

Next, we then apply the AT condition to the AoI minimization problem with both stochastic arrivals and unreliable channels, and for the first time in the literature, affirmatively establish its Whittle indexability. Due to the 2-dimensional state-space for the stochastic-arrival setting, even applying the AT condition is quite involved (as we cannot easily relate to the recurrent state as in the generate-at-will setting). Nonetheless, the intuition behind the AT condition allows us to develop a novel coupling approach, which successfully verifies the AT condition for this setting. This coupling approach may be of independent interest to other problems. The structural property of the AoI minimization problem and the near-optimality of the Whittle index policy are verified through our numerical results, where we also introduce an improved version of the algorithm in [20] and [21] for fast computation of the Whittle index.

The rest of the paper is structured as follows. In Section 2, we introduce both the AoI minimization problem for the stochastic-arrival setting and a more general RMAB formulation, and define the Whittle index. In Section 3, we propose our active-time (AT) condition for proving Whittle indexability for the general RMAB formulation with finite state-space, which can potentially be applicable to many problem settings. In Section 4, we apply our AT condition to the AoI minimization problem specifically, and use novel coupling approaches to prove its Whittle indexability, for both the generate-at-will and the stochastic-arrival settings. Numerical results are presented in Section 5, and then we conclude.

II. System Model

Below, we first introduce the system model for the AoI minimization problem with stochastic arrivals and unreliable channels, which becomes a special case of a more general RMAB model that we present afterwards. Some of our key results, such as the AT condition, apply to the general model.

Consider a base-station (BS) that transmits information updates to N user-equipments (UEs) in the downlink. Assume that time is slotted, such that each transmission takes exactly one unit of time. The BS has channel capacity C, i.e., it can transmit to at most C UEs at each time slot. We assume that $C < N$. Assume that there are N buffers at the BS, each corresponds to a UE. The information update packets for UE n arrive to the BS following a Bernoulli process with rate $p_g(n) \in (0,1]$, independently across the UEs. Due to the uncertainty of the wireless network, each packet transmission from the BS to UE n succeeds with probability $p_s(n) \in (0,1]$, independently of other transmissions. Let $u_n(t) \in \{0,1\}$ denote the scheduling decision for UE n at time t. That is, $u_n(t) = 1$ denotes that the BS schedules a transmission to UE n at time t, and $u_n(t) = 0$ otherwise. Thus, we must have $\sum_n u_n(t) \le C$ for all time t. Define $h'_n(t)$ to be the AoI of UE n at time t, which is the time elapsed from the generation time of the last received packet at UE n to the current time t. Define $a'_n(t)$ to be the time elapsed from the arrival time of the newest packet in the buffer for UE n at the BS to the current time t. Without loss of generality, we take the convention that, at each time slot, the scheduled transmission happens before the new arrival at that time slot. As a consequence, a successful transmission to UE n at time t will reduce the AoI from $h'_n(t)$ to $a'_n(t)$. Let $d'_n(t) = h'_n(t) - a'_n(t)$, which can be viewed as the improvement in AoI if a successful transmission to UE n takes place. Thus, we can describe the AoI evolution of UE n by the evolution of $a'_n(t)$ and $d'_n(t)$ as:

$$a'_n(t+1) = \begin{cases} 1 & \text{new arrival to } n\text{'s buffer}, \\ a'_n(t)+1 & \text{otherwise}, \end{cases} \tag{1}$$

$$d'_n(t+1) = \begin{cases} a'_n(t) & \text{transmission success and new arrival}, \\ 0 & \text{transmission success and no new arrival}, \\ a'_n(t)+d'_n(t) & \text{new arrival and no transmission success}, \\ d'_n(t) & \text{otherwise}. \end{cases} \tag{2}$$

For technical reasons, our main result on the AT condition requires a finite state space. Thus, we approximate the true state-evolution by a state truncation. Specifically, define the operator $\wedge$ such that $x \wedge y = \min\{x, y\}$. Given two arbitrarily large integers $K_a, K_d \in \mathbb{N}_+$, we define $a_n(t) = a'_n(t) \wedge K_a$ and $d_n(t) = d'_n(t) \wedge K_d$. In the rest of the paper, we will take the tuple $s_n(t) = \big(a_n(t), d_n(t)\big) \in \{1, ..., K_a\} \times \{0, 1, ..., K_d\}$ as the state of UE n at time t. Intuitively, when K_a and K_d are large, the inaccuracy due to this state truncation will vanish.

We can now formulate the AoI minimization problem for the above system as a Restless Multi-Armed Bandit (RMAB) problem, where each arm denotes a UE. Let $\bar{S}(t) = \big\{\big(a_1(t), d_1(t)\big), ..., \big(a_N(t), d_N(t)\big)\big\}$ be the system state of all arms at time t and denote the action as $\bar{U}(t) = \{u_1(t), ..., u_N(t)\} \in \bar{\mathcal{A}} \triangleq \{0,1\}^N$. Let policy $\bar{\pi}$ be a function mapping each state $\bar{S}$ to an action $\bar{U}$. Suppose that the current state $\big(a_n(t), d_n(t)\big)$ for UE n is (a, d). Then, the state transition probabilities for UE n under the passive action, i.e.,

$u_n(t) = 0$, are

$$\begin{aligned}
&\mathbf{Pr}\{(a,d) \to \big((a+1)\wedge K_a, d\big) | u_n(t)=0\} = 1 - p_g(n) \\
&\mathbf{Pr}\{(a,d) \to \big(1, (a+d)\wedge K_d\big) | u_n(t)=0\} = p_g(n),
\end{aligned} \tag{3}$$

and that under the active action, i.e., $u_n(t) = 1$, is

$$\begin{aligned}
&\mathbf{Pr}\{(a,d) \to \big((a+1)\wedge K_a, d\big) | u_n(t)=1\} = \\
&\qquad\qquad \big(1-p_g(n)\big)\big(1-p_s(n)\big) \\
&\mathbf{Pr}\{(a,d) \to \big(1, (a+d)\wedge K_d\big) | u_n(t)=1\} = \\
&\qquad\qquad p_g(n)\big(1-p_s(n)\big) \\
&\mathbf{Pr}\{(a,d) \to \big((a+1)\wedge K_a, 0\big) | u_n(t)=1\} = \\
&\qquad\qquad \big(1-p_g(n)\big)p_s(n) \\
&\mathbf{Pr}\{(a,d) \to \big(1, a\wedge K_d\big) | u_n(t)=1\} = p_g(n)p_s(n).
\end{aligned} \tag{4}$$

Let $v(\cdot) : \mathbb{N} \to \mathbb{R}$ be a strictly increasing function representing the penalty cost on the AoI value. Let $\delta_n(t) = 1$ if the transmission to UE n at time t is successful, and $\delta_n(t) = 0$ if such a transmission fails (or if there is no such transmission). Let $\beta \in (0,1)$ be a discount factor. Our objective is to minimize the discounted expected total cost of all arms that starting at an arbitrary initial state $\bar{S}_0$ [12], i.e.,

$$\begin{aligned}
\min_{\bar{\pi}} \sum_{t=0}^{\infty} \beta^t \sum_{n=1}^{N} \mathbb{E}_{\bar{S}_0}^{\bar{\pi}} \Big[& v\big(a_n(t) + d_n(t) \\
& \times \big(1 - u_n(t)\delta_n(t)\big)\big)\Big]
\end{aligned} \tag{5}$$

$$\text{s.t.} \sum_{n=1}^{N} u_n(t) \le C, \text{for all time } t, \tag{5a}$$

where $\mathbb{E}_{\bar{S}_0}^{\bar{\pi}}[\cdot]$ denote the expectation under policy $\bar{\pi}$ starting from the initial state $\bar{S}_0$.

At this point, it should also be clear that our AoI minimization problem is a special case of a more general RMAB problem, where we can replace the state $\big(a_n(t), d_n(t)\big)$ by a general state $s_n(t)$, replace the AoI penalty in (5) by a general per-stage cost $c\big(s_n(t), u_n(t)\big)$, and replace (3) and (4) by arbitrary state transition laws P^0 and P^1. The discussion below and the results of Section 3 will be applicable to this general RMAB problem.

Unfortunately, such an RMAB problem is known to suffer from the curse-of-dimensionality, i.e., the complexity grows exponentially with N. As in [22], we follow Whittle's approach and instead study a relaxed problem, where the decision for each UE can be decomposed. Specifically, we first relax the hard constraint (5a) to a soft constrain in the discounted-sum form, i.e.,

$$\min_{\bar{\pi}} \sum_{t=0}^{\infty} \beta^t \sum_{n=1}^{N} \mathbb{E}_{\bar{S}_0}^{\bar{\pi}} \left[c\big(s_n(t), u_n(t)\big) \right] \tag{6}$$

$$\text{s.t.} \sum_{t=0}^{\infty} \beta^t \sum_{n=1}^{N} \mathbb{E}_{\bar{S}_0}^{\bar{\pi}} \left[u_n(t) \right] \le \frac{C}{1-\beta}. \tag{6a}$$

Since (5a) implies (6a), the solution to (6) becomes a lower bound to that of (5). Then, we associate a dual variable λ to (6a) and apply Lagrange relaxation. Minimization of the corresponding Lagrangian can be decomposed into N per-UE problems, where the objective of UE n is

$$\min \sum_{t=0}^{\infty} \beta^t \mathbb{E}\big[c\big(s_n(t), u_n(t)\big) + \lambda u_n(t) \big]. \tag{7}$$

If we interpret λ as an activation price of being scheduled, in (7) each UE n simply minimizes its own per-stage cost plus the activation cost (both of which are discounted in time). This per-UE problem can also be viewed as an MDP, but with a much smaller state space $\mathcal{S}_n = \{1,...,K_a\} \times \{0,1,...,K_d\}$ and a much smaller action space $\mathcal{A}_n = \{0,1\}$.

We can now define Whittle indexability based on this per-UE MDP. Let π_n be the policy of UE n mapping $\mathcal{S}_n$ to $\mathcal{A}_n$. Given λ, we call a policy that achieves the minimum in (7) as the optimal policy, denoted as $\pi_{n\lambda}^*$. Define the passive set to be $\mathcal{P}_n(\lambda) = \big\{ s_n \in \mathcal{S}_n \big| \text{There exists } \pi_{n\lambda}^* \text{ such that } \pi_{n\lambda}^*(s_n) = 0 \big\}$, i.e., the set of states that will take the passive action under at least one optimal policy. Note that this definition allows the possibility that the optimal policy might not be unique. The definition of Whittle indexability [22] is as follows.

Definition 1. *The per-UE problem* (7) *is Whittle indexable if its MDP satisfies the following properties:*

1) $\mathcal{P}_n(+\infty) = \mathcal{S}_n$ *and* $\mathcal{P}_n(-\infty) = \emptyset$*,*
2) $\mathcal{P}_n(\lambda_1) \subseteq \mathcal{P}_n(\lambda_2)$*, for all* $\lambda_1 < \lambda_2$.

If Definition 1 holds, we can define the Whittle index as

$$w(s_n) = \inf \big\{ \lambda | s_n \in \mathcal{P}_n(\lambda) \big\}, \tag{8}$$

which intuitively is the activation cost λ such that the active and passive actions break even.

Once the Whittle index for every state is calculated, the Whittle index policy will simply activate arms with the highest Whittle indices. Further, such a policy has been shown to be asymptotically optimal when the system size is large [15], [16]. However, as we discussed earlier, one of the main challenges for Whittle index policy is to prove the indexability of the per-UE problem. This is particularly challenging when the value function of the per-UE problem (and thus the optimal policy) cannot be derived in closed form. In the next section, we will introduce our AT condition, which enables us to prove Whittle indexability without full knowledge of the value function and the optimal policy.

III. The Active Time (AT) Condition

Recall that at the end of Section 2, we have decoupled a general RMAB problem into per-UE sub-problems through the Whittle relaxation (6). In the rest of the paper, we will focus on this per-UE sub-problem. Thus, for ease of exposition, we will omit the UE index n when there is no source of confusion. This per-UE problem then has the state s, the action u and the policy π in the corresponding state space $\mathcal{S}$, action space $\mathcal{A}$ and policy space Π. Let $K = |\mathcal{S}|$ denote the size of state space $\mathcal{S}$ and λ denote the activation cost. The per-stage cost function is given by $c(s,u) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and let $c_\lambda(s,u) = c(s,u) + \lambda u$. The state transition matrix under action u is denoted as P^u.

Let $V_\lambda^\pi(s)$ be the value function for state s, under policy π and given the activation cost λ of the per-UE MDP. Then, we can write

$$V_\lambda^\pi(s) = C_s^\pi + \lambda T_s^\pi, \tag{9}$$

where

$$T_s^\pi = \mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \beta^t u_t^\pi \Big| s_0 = s\Big] \tag{10}$$

is the expected discounted total activation time and

$$C_s^\pi = \mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \beta^t c(s_t, u_t^\pi) | s_0 = s\Big] \tag{11}$$

is the expected discounted total cost under policy π and starting from state s. Note that u_t^π is the chosen action at time t by policy π. Similarly, let $Q_\lambda^\pi(s,u)$ be the action-value function for state s and action u, under policy π and given the activation cost λ of the per-UE MDP. Then, we can also write

$$Q_\lambda^\pi(s,u) = C_s^\pi(u) + \lambda T_s^\pi(u), \tag{12}$$

where

$$T_s^\pi(u) = \mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \beta^t u_t^\pi \Big| s_0 = s, u_0^\pi = u\Big], \tag{13}$$

$$C_s^\pi(u) = \mathbb{E}_\pi\Big[\sum_{t=0}^{\infty} \beta^t c(s_t, u_t^\pi) | s_0 = s, u_0^\pi = u\Big].$$

Obviously, both $V_\lambda^\pi(s)$ and $Q_\lambda^\pi(s,u)$ are linear functions in λ with slopes T_s^π and $T_s^\pi(u)$, respectively.

Let $V_\lambda^*(s)$ be the optimal value function and $Q_\lambda^*(s,u)$ be the optimal action-value function. By definition,

$$\begin{aligned} V_\lambda^*(s) &= \min_{\pi\in\Pi} V_\lambda^\pi(s) \\ Q_\lambda^*(s,u) &= \min_{\pi\in\Pi} Q_\lambda^\pi(s,u) \end{aligned} \tag{14}$$

Consider a smaller subset of policies $\Pi' \subseteq \Pi$, such that Π' contains every possible optimal policy[1]. Then, (14) becomes

$$V_\lambda^*(s) = \min_{\pi\in\Pi'} V_\lambda^\pi(s) \tag{15a}$$

$$Q_\lambda^*(s,u) = \min_{\pi\in\Pi'} Q_\lambda^\pi(s,u). \tag{15b}$$

The following lemma has been shown in earlier work, e.g. [20]. We provide the proof below for the sake of completeness.

Lemma 2. *Both $V_\lambda^*(s)$ and $Q_\lambda^*(s,u)$ are continuous, monotonically increasing, concave and piece-wise linear functions with respect to λ.*

Proof. First, being the point-wise minimum of many linear functions in λ, both $V_\lambda^*(s)$ and $Q_\lambda^*(s,u)$ must be continuous and concave in λ. Using the fact that $T_s^\pi \geq 0$ and $T_s^\pi(u) \geq 0$, we conclude that both functions are monotonically increasing. Since the state space is finite, the policy space Π is also finite. Thus, the minimum of a finite number of linear functions produces a piece-wise linear function. □

[1]For example, for the AoI minimization problem that we studied ealier, one can show that the optimal policy must be a threshold type. Then, Π' can be taken to only include such policies.

Theorem 3. *(The AT condition) The per-UE MDP is Whittle indexable if it satisfies the following conditions:*

1) $\mathcal{P}(+\infty) = \mathcal{S}$ *and* $\mathcal{P}(-\infty) = \emptyset$.
2) $T_s^\pi(1) \geq T_s^\pi(0)$, *for all state* $s \in \mathcal{S}$ *and for all* $\pi \in \Pi'$, *where* Π' *is a subset of* Π *that contains all possible optimal policies.*

Proof. Define $\Delta Q_s^*(\lambda) = Q_\lambda^*(s,1) - Q_\lambda^*(s,0)$. By Lemma 2, $\Delta Q_s^*(\lambda)$ is also a continuous and piece-wise linear function. Since $Q_\lambda^*(s,1)$ and $Q_\lambda^*(s,0)$ are piece-wise linear, their right derivatives are always well defined. Specifically, for any given λ, define $D(u,\lambda) = \operatorname{argmin}_{\pi\in\Pi'} Q_\lambda^\pi(s,u)$ and

$$\tau(u,\lambda) = \min\{T_s^\pi(u) : \pi \in D(u,\lambda)\}. \tag{16}$$

Then, we have $\frac{d^+ \Delta Q_s^*}{d\lambda^+} = \tau(1,\lambda) - \tau(0,\lambda)$.

To prove the result of the theorem, it suffices to show the following statement: for all λ, there exists $\pi^\lambda \in D(0,\lambda) \cap D(1,\lambda)$ such that $\tau(1,\lambda) - \tau(0,\lambda) = T_s^{\pi^\lambda}(1) - T_s^{\pi^\lambda}(0)$. To see why this statement is sufficient, since $D(0,\lambda) \cap D(1,\lambda) \subseteq \Pi'$, applying the second condition of the theorem on policy π^λ, we conclude that $\Delta Q_s^*(\lambda)$ must be a monotonically increasing function at all λ. Now, suppose that $s \in \mathcal{P}(\lambda_1)$, i.e., some optimal policy will take the passive decision at state s under λ_1. By the definition of the action-value function, it implies that $Q_\lambda^*(s,0) \leq Q_\lambda^*(s,1)$ and $\Delta Q_s^*(\lambda_1) \geq 0$. Then, for any $\lambda_2 > \lambda_1$, we must have, from the monotonicity, $\Delta Q_s^*(\lambda_2) \geq 0$. Hence, we must have $s \in \mathcal{P}(\lambda_2)$. Combined with the first condition of the theorem, we have proven Whittle indexability.

It remains to show the statement that, for any λ, there exists $\pi^\lambda \in D(0,\lambda) \cap D(1,\lambda)$ such that $\tau(1,\lambda) - \tau(0,\lambda) = T_s^{\pi^\lambda}(1) - T_s^{\pi^\lambda}(0)$. First, by the property of MDP, the optimal policy for an MDP should simultaneously minimize the value function and the action-value function for all states and actions. Further, by the piece-wise linearity of $V_{\lambda'}^*(s)$ as a function of λ', there must exists a policy π^1 that attains the minimum in (15a) for all states and all λ' in an interval $[\lambda, \lambda+\delta]$ for some $\delta > 0$. The policy π^1 will then also minimize $Q_{\lambda'}^*(s,0)$ and $Q_{\lambda'}^*(s,1)$ for all states and actions, and over all $\lambda' \in [\lambda, \lambda+\delta]$. Thus, π^1 must attain the minimum in (16) at λ for both $u=0$ and $u=1$, i.e., $T_s^{\pi^1}(0) = \tau(0,\lambda)$ and $T_s^{\pi^1}(1) = \tau(1,\lambda)$. The theorem then follows. □

The significance of Theorem 3 is that it provides a much easier way to verify Whittle indexablity, by simply comparing the expected total active times at different initial actions. Interestingly, in his original paper [22], Whittle also stated a similar sufficient condition for indexability that $T_s^\pi(1) - T_s^\pi(0) \geq 0$ for the *optimal policy* π given λ (see page 296 of [22]). One could then argue that, since the set Π' contains all possible optimal policies, our condition 2 of Theorem 3 (that holds for all policies in Π') implies Whittle's sufficient condition (that holds for the optimal policy π at λ). However, Whittle did not provide a rigorous proof for his condition. Specifically, Whittle started his argument by stating that the derivative of

$\Delta Q_s^*(\lambda)$ is equal to $T_s^\pi(1) - T_s^\pi(0)$ for the optimal policy π at λ. Unfortunately, this statement is not always true, especially when there are multiple optimal policies at λ (See (16)). Instead, our proof of Theorem 3 correctly handles such cases, by finding the right policy π^λ that produces the correct derivative for $\Delta Q_s^*(\lambda)$. In practice, verifying condition 2 of the AT condition for all policies in Π' is also easier since we do not need to know the exact optimal policy.

Note that Theorem 3 is also closely related to GCL [17] and PCL [18]; both of which involve conditions on $T_s^\pi(1)$ and $T_s^\pi(0)$. When we take Π' to be the set Π of all possible policies, our condition then become very similar to GCL. Both our condition and the PCL condition are more powerful than GCL because we only need to verify $T_s^\pi(1) > T_s^\pi(0)$ for a much smaller set Π' of policies. However, the AT condition is even easier and more powerful than PCL for the following reasons:

1) The PCL condition requires one to verify an additional admissibility condition [18], which seems non-trivial. The AT condition does not require this additional step.
2) The PCL condition requires one to pick a class $\mathcal{F}$ of state subsets that satisfies certain properties. Our AT condition replace it by Π' that contains all optimal policies, which is much easier to understand and work with.
3) The PCL condition requires $T_s^\pi(1) > T_s^\pi(0)$. As we have shown in Theorem 3, it turns out that $T_s^\pi(1) \geq T_s^\pi(0)$ is sufficient according to our AT condition.

Therefore, we expect that our AT condition can potentially be applied to a larger set of problems.

IV. Indexability of the AoI minimization problem

In this section, we return to the more-specific AoI minimization problem defined at the beginning of Section 2, and we will show how to apply our AT condition to prove its Whittle indexability. As in Section 3, since we focus on a per-UE MDP, we will omit the index n for UE when there is no source of confusion. Note that when $p_g = 1$, our model in Section 2 reduces to the generate-at-will model, where the Whittle indexability has already been shown by deriving an closed-form expression of the value function [5]. We will first revisit this generate-at-will setting to illustrate the simplicity of applying the AT condition, i.e, we will show how to establish indexability without deriving the value function. Then, we will move on to the more challenging stochastic-arrival case where $p_g < 1$, whose indexability has not been established before.

A. The Generate-at-Will Setting

In the generate-at-will setting, the value of a is always 1. Thus, we can simplify the state into 1-D. Recall the operator $\wedge$ and the state truncation threshold K_d defined in Section 2. Define $h_n(t) = h'_n(t) \wedge K_d$. We omit the UE index n and denote the AoI as $h(t)$, which we will use as the 1-D state. Then, the state evolution can be described by

$$h(t+1) = \begin{cases} 1 & \text{transimission success} \\ \big(h(t)+1\big) \wedge K_d & \text{otherwise,} \end{cases} \tag{17}$$

and the Bellman equation is

$$\begin{aligned} V_\lambda^*(h) =& \min\big\{Q_\lambda^*(h,0), Q_\lambda^*(h,1)\big\}, \\ Q_\lambda^*(h,0) =& v(h) + \beta V_\lambda^*\big((h+1)\wedge K_d\big), \\ Q_\lambda^*(h,1) =& \lambda + p_s\big[v(0) + \beta V_\lambda^*(1)\big] \\ &+ (1-p_s)\big[v(h) + \beta V_\lambda^*\big((h+1)\wedge K_d\big)\big], \end{aligned} \tag{18}$$

where $V_\lambda^*(h)$ is the value function for state h and $Q_\lambda^*(h,u)$ is the action-value function for state h and action u, as defined earlier in Section 3. The state space is $\mathcal{S} = \{1, ..., K_d\}$.

Note that state 1 is the recurrent state: when there is a successful transmission, every state will return to state 1. We next show some properties of this per-agent MDP.

Lemma 4. *$V_\lambda^*(h)$ is monotonically increasing in h.*

Lemma 5. *For each λ, there exists a threshold $H_\lambda \in \{1, ..., K_d+1\}$, such that the passive set, denoted as $\mathcal{P}(\lambda)$, is given by $\mathcal{P}(\lambda) = \{h : h < H_\lambda \text{ and } h \in \mathcal{S}\}$.*

Here we omit the proof because similar results have been developed in earlier work, e.g., [5].

We next show that the MDP satisfies the two conditions in Theorem 3. Define Π' to be the set of policies that are of a threshold type as stated in Lemma 5, i.e., each policy in Π' will take passive actions on $\{h : h < H\}$ for some H, and take active actions otherwise. By varying H, Π' will include all possible optimal policies according to Lemma 5.

Theorem 6. *$T_h^\pi(1) \geq T_h^\pi(0)$, for all $\pi \in \Pi'$ and $h \in \mathcal{S}$.*

Proof. Let us first define two Markov chains, chain 1 and chain 0. The two chains have exactly the same initial state $s_0 = h$, activation cost λ and state transition probability. Further, they both follow the same policy $\pi \in \Pi'$ after time 1. However, the difference is at time 0, where chain 1 uses the active action $u = 1$, and chain 0 uses the passive action $u = 0$. We will then compare their active times. To ease our analysis, we couple the two chains in the following way. From time 0 onwards, the j-th transmission (i.e., active action) of chain 1 will have the same channel success/failure event as the j-th transmission of chain 0. Note that this coupling is feasible because it does not alter the state transition probability, i.e., both chains still have the same channel success probability p_s.

Denote the time of the first *successful* transmission of chain 1 and chain 0 as $\tau(1)$ and $\tau(0)$, respectively. To show the theorem, it suffices to show the following two inequalities:

$$\mathbb{E}\Big[\sum_{t=0}^{\tau(1)} \beta^t u_1^t\Big] \geq \mathbb{E}\Big[\sum_{t=0}^{\tau(0)} \beta^t u_0^t\Big], \tag{19}$$

$$\mathbb{E}\Big[\sum_{t=\tau(1)+1}^{\infty} \beta^t u_1^t\Big] \geq \mathbb{E}\Big[\sum_{t=\tau(0)+1}^{\infty} \beta^t u_0^t\Big], \tag{20}$$

where u_1^t and $u_0^t \in \{0,1\}$ are the actions chosen by chain 1 and chain 0, respectively, at time t and $\beta \in (0,1)$. Indeed, if we sum up (19) and (20), it exactly gives $T_h^\pi(1) \geq T_h^\pi(0)$ by (13), and the theorem then follows.

We next show (19). Note that $\pi \in \Pi'$ satisfies the threshold structure in Lemma 5. Thus, as long as a chain's state is above the threshold H, it will keep transmitting until success. Suppose that at time $t' \geq 1$, the state of chain 0 reaches the threshold H. Recall that chain 1 uses the active action at time 0. We then divide into two cases.

Case 1: If chain 1 has a successful transmission at time $t = 0$. By our coupling, chain 0 should also have a successful transmission at the first time that it transmits, which occurs at time $t' \geq 1$. We then have $\tau(1) = 0$ and $\tau(0) = t'$. Obviously, the inequality (19) holds.

Case 2: If chain 1 has a failure at time $t = 0$. By our coupling, the two chains will then have the same state evolution until a channel success occurs. In particular, from time 1 to $t' - 1$, both chains will take passive actions. At and after t', both chains will take active actions, until each of their first successful transmission. By our coupling, since chain 1 has an extra attempt at $t = 0$, it must succeed by exactly 1 time-slot earlier than chain 0, i.e., $\tau(1) = \tau(0) - 1$. Since $\beta^0 u_1^0 = 1 > \beta^{\tau(0)} u_0^{\tau(0)}$, the inequality (19) holds.

In summary, (19) holds in both cases.

We finally show (20). By the above proof of (19), we have also shown that $\tau(1) < \tau(0)$. Note that both chains return to the same state 1 after their first successful transmission. Further, since they follow the same policy and have the same channel events by our coupling, the future state-evolution of the two chains are exactly the same, except for the offset $\tau(1)$ and $\tau(0)$ in time. Then, we must have (20) and the result of the theorem then follows. □

By (18), Lemma 4 and Lemma 5, it is easy to verify that $H_\lambda = K_d + 1$ when $\lambda = +\infty$ and $H_\lambda = 1$ when $\lambda = -\infty$, where the passive sets are $\mathcal{S}$ and $\emptyset$, respectively. Combining with Theorem 6, we conclude that the per-UE MDP is Whittle indexable. Since our analysis does not depends on K_d, the indexability holds for arbitrarily large K_d.

B. The Stochastic-Arrival Setting

We now consider the more-challenging stochastic-arrival setting with packet generation rate $p_g < 1$. In this case, we have to use (a, d) as the 2-D system state, and the state space is $\mathcal{S} = \{1, ..., K_a\} \times \{0, ..., K_d\}$. Using (3) and (4), we can express the Bellman equation as

$$\begin{aligned} V_\lambda^*(a,d) &= \min\big\{Q_\lambda^*(a,d,0), Q_\lambda^*(a,d,1)\big\}, \\ Q_\lambda^*(a,d,0) &= v(a+d) + p_g \beta V_\lambda^*\big(1, (a+d)\wedge K_d\big) \\ &\quad + (1-p_g)\beta V_\lambda^*\big((a+1)\wedge K_a, d\big), \\ Q_\lambda^*(a,d,1) &= (1-p_s)\Big[v(a+d) + p_g\beta V_\lambda^*\big(1,(a+d)\wedge K_d\big) \\ &\quad + (1-p_g)\beta V_\lambda^*\big((a+1)\wedge K_a, d\big)\Big] \\ &\quad + p_s\Big[v(a) + p_g\beta V_\lambda^*(1, a\wedge K_d) \\ &\quad + (1-p_g)\beta V_\lambda^*\big((a+1)\wedge K_a, 0\big)\Big] + \lambda. \end{aligned} \tag{21}$$

Compared to the generate-at-will setting, the state evolution and the Bellman equation become much more complicated. However, some properties of the optimal policy can also be found, similar to Lemma 4 and Lemma 5.

Lemma 7. *At any fixed state a, $V_\lambda^*(a,d)$ is a monotonically increasing function with respect to d.*

Proof. The proof utilizes [23, Prop. 3.1]. Let $\xi_t \in \Xi$ be the discrete-time random process that represents the random channel event and packet generation event. Then, we can describe the state transitions by a deterministic function f taking ξ_t as an argument, i.e., $(a_{t+1}, d_{t+1}) = f\big((a_t, d_t), u_t, \xi_t\big)$. We then introduce the following partial order on the state space. For two states with the same values of a, we define $(a, d_1) > (a, d_2)$ if $d_1 > d_2$. For two states with different values of a, they are not comparable. Then, for $(a, d_1), (a, d_2) \in \mathcal{S}$ with $d_1 > d_2$, we can verify that:

1) $f\big((a,d_1),u,\xi\big) \geq f\big((a,d_2),u,\xi\big)$, which is a result of the state transition laws (3) and (4).
2) $c_\lambda((a,d_1),u) \geq c_\lambda((a,d_2),u)$, since $v(\cdot)$ is a strictly increasing function.
3) ξ_t, which describes the channel event and packet arrival event, is independent of the current state s_t.

The author of [23] show that, if these conditions hold, then the value function must be increasing in the partial order of the states that we defined, i.e., in the d dimension. The result of the lemma then follows. □

Lemma 8. *There exists a threshold $H_\lambda(a) \in \{0, ..., K_d + 1\}$ for each a, such that the passive set, denoted as $\mathcal{P}(\lambda)$, satisfies that $\mathcal{P}(\lambda) = \{(a,d) : d < H_\lambda(a)$ and $(a,d) \in \mathcal{S}\}$*

Proof. We first define $\mu(a,d) = Q_\lambda^*(a,d,1) - Q_\lambda^*(a,d,0)$, i.e., the difference between the action-value functions with the active and passive initial actions. By (21), we have

$$\begin{aligned} \mu(a,d) =& \lambda - p_s\Big\{v(a+d) - v(a) \\ &+ p_g\beta\Big[V_\lambda^*\big(1,(a+d)\wedge K_d\big) - V_\lambda^*\big(1, a\wedge K_d\big)\Big] \\ &+ (1-p_g)\beta\Big[V_\lambda^*\big((a+1)\wedge K_a, d\big) \\ &\quad - V_\lambda^*\big((a+1)\wedge K_a, 0\big)\Big]\Big\} \end{aligned} \tag{22}$$

Since $v(\cdot)$ is a strictly increasing function and $V_\lambda^*(a,d)$ is monotonically increasing in d, we have that $\mu(a,d)$ is strictly decreasing in d. Suppose that $(a,d) \notin \mathcal{P}(\lambda)$. Then every optimal policy will take the active action at state (a,d), which implies that $\mu(a,d) < 0$. Then, by the earlier argument, for all $d^+ > d$, we must have $\mu(a,d^+) < 0$, and hence $(a,d^+) \notin \mathcal{P}(\lambda)$. The lemma then follows. □

Thanks to Lemma 8, we can then choose the set Π' of polices as those that satisfy the threshold structure in Lemma 8, i.e., each policy in Π' will take passive actions over the set $\{(a,d) : d < H(a)\}$, where $H(a), a = 1, 2, ..., K_a$ are the thresholds. It remains to show that condition 2 of Theorem 3 holds for any policy π in Π'. Unfortunately, the verification of condition 2 is much more challenging when the state space is

2-D, because the chain does not return to a common state after each successful transmission. Instead, it returns to multiple possible states with different values of a. To overcome this difficulty, below we use a novel coupling analysis.

Similar to the proof of Theorem 6, we define two Markov chains, chain 1 and chain 0, with exactly the same initial state $s_0 = s$, activation cost λ and the state transition probability. They both follow the same policy $\pi \in \Pi'$ after time 1, but have different initial actions at $t = 0$, where chain 1 takes $u^0 = 1$ and chain 0 takes $u^0 = 0$. We apply the same channel event coupling as in the proof of Theorem 6: From time 0 onwards, the j-th transmission (i.e., active action) of chain 1 will have the same channel success/failure event as the j-th transmission of chain 0. In addition, we also couple the arrivals of the two chains in the following way: the arrival event of chain 1 at time t must be the same as the arrival event of chain 0 at time t. Note that both of the two couplings do not change the state transition probability because both chains have the same channel success probability p_s and packet generation rate p_g.

Since both chains start with the same state and the evolution of the a-component of the state is independent of channel events, our coupling on the arrival events ensures that the a-component of the states of the two chains must be equal at all time. Thus, we can focus our attention on the difference in the d-component. This gives us some hope that we can borrow from the 1-D analysis in the generate-at-will case. However, there is still significant difference in how the d-component evolves when the state space is 2-D.

(1) In the generate-at-will setting, the state h increases by 1 every time slot until a successful transmission, in which case the state h reduces to 1. In contrast, in the stochastic-arrival setting, the next state depends on not only the channel success, but also the packet generation event.

(2) In the generate-at-will setting, the threshold is unique and fixed since the state space is 1-D. In contrast, the threshold on d in the stochastic-arrival setting depends on a.

We therefore need new techniques to analyze the evolution of the d-component after a sequence of channel failures and successes. Towards this end, define $\tau_u(j)$ to be the occurrence time of the j-th successful transmission for chain $u = 0, 1$. Define $d_u(t)$ to be the value of the d-component at time t for chain u.

We next show two simple consequences due to the threshold policy $\pi \in \Pi'$ and the state evolution.

Lemma 9. *(Two key properties)*

(1) For any time interval $[t_1, t_2)$, if $d_1(t) \geq d_0(t)$ holds for all $t \in [t_1, t_2)$, then whenever chain 0 has an active action at time $t \in [t_1, t_2)$, chain 1 must also have an active action at the same time.

(2) Suppose that chain 0 have a channel success at t_1. Further, after t_1, suppose that chain 1 does not have a channel success until t_2. Then, we have $d_1(t) \geq d_0(t)$ for $t \in [t_1, t_2)$.

Proof. To prove part (1), recall that through our coupling, the a-component of the two chains are always the same. If chain 0 have an active action at time $t \in [t_1, t_2)$, then $d_0(t)$ must be above the a-dependent threshold $H(a)$ at that time. Since $d_1(t) \geq d_0(t)$, $d_1(t)$ must also be above the threshold $H(a)$, and hence chain 1 must also take the active action.

To prove part (2), recall again that through our coupling, the a-component of the two chains are always the same. Suppose at time $t_1 - 1$, both chains are at state a. By the state transition law in (3) and (4), once chain 0 successfully transmits at t_1, we have $d_0(t_1) = 0$ (if there is no arrival) or $d_0(t_1) = a \wedge K_d$ (if there is an arrival). If chain 1 also successfully transmits at t_1, then $d_1(t_1) = d_0(t_1)$ by our coupling. Otherwise, we have $d_1(t_1) = d_1(t_1 - 1)$ (if there is no arrival) or $d_1(t_1) = (d_1(t_1 - 1) + a) \wedge K_d$ (if there is an arrival). In both cases, we must have $d_1(t_1) \geq d_0(t_1)$ because $d_1(t_1 - 1) \geq 0$. After time t_1, we must also have $d_1(t_1) \geq d_0(t_1)$ until chain 1 have a transmission success. Thus, part (2) of the lemma follows. □

We now state a key lemma on $\tau_1(j)$ and $\tau_0(j)$, which implies that the successful transmission times of chain 1 are always ahead of the respective ones of chain 0.

Lemma 10. *$\tau_0(j) \geq \tau_1(j)$ holds for all $j \in \mathbb{N}_+$.*

Proof. We prove by induction. First, we check the initial case where $j = 1$. Our objective is to show $\tau_0(1) \geq \tau_1(1)$. Recall that chain 1 uses the active action at time $t = 0$. Similar to the proof of Theorem 6, we can then divide into 2 cases.

Case 1: If chain 1 has a successful transmission at time $t = 0$. By our channel-event coupling, chain 0 should also have a successful transmission at the first time it transmits, which occurs at some time $t > 0$. Then, $\tau_0(1) > \tau_1(1)$ holds.

Case 2: If chain 1 has a failure at time $t = 0$. The two chains will then have the same state evolution until a success occurs in either one of them. However, by our channel-event coupling, chain 1 must succeed earlier since it has an extra active action at time $t = 0$, which implies that $\tau_0(1) > \tau_1(1)$.

Now we prove the induction step. We assume that $\tau_0(j - 1) \geq \tau_1(j - 1)$, and we wish to show that $\tau_0(j) \geq \tau_1(j)$. Again, divide into 2 cases.

Case 1: If $\tau_0(j - 1) \geq \tau_1(j)$, then $\tau_0(j) \geq \tau_1(j)$ trivially hold because $\tau_0(j) \geq \tau_0(j - 1)$.

Case 2: If $\tau_0(j - 1) < \tau_1(j)$. We proceed by contradiction. Assume in contrary that $\tau_0(j) < \tau_1(j)$. By property (2) in Lemma 9, we have $d_1(t) \geq d_0(t)$ for $t \in \big[\tau_0(j - 1), \tau_0(j)\big)$. Thus, by property (1) in Lemma 9, for each active action of chain 0 in the time interval $\big[\tau_0(j-1), \tau_0(j)\big)$, there is an active action in chain 1 at the same time. Further, chain 1 may take additional active actions in the interval $\big[\tau_0(j-1), \tau_0(j)\big)$. Thus, the number of active actions that chain 1 takes from $\tau_1(j-1)$ to $\tau_0(j)$ is always greater than or equal to the number of active actions that chain 0 takes from $\tau_0(j - 1)$ to $\tau_0(j)$. However, by our channel-event coupling, starting from their respective $(j - 1)$-th successful transmission, their sequences of channel successes and failures are identical. Therefore, since chain 0 succeeds at time $\tau_0(j)$, chain 1 must have succeeded at or before time $\tau_0(j)$. This would have implied $\tau_0(j) \geq \tau_1(j)$, which contradicts our initial assumption that $\tau_0(j) < \tau_1(j)$. Therefore, we must have $\tau_0(j) \geq \tau_1(j)$. □

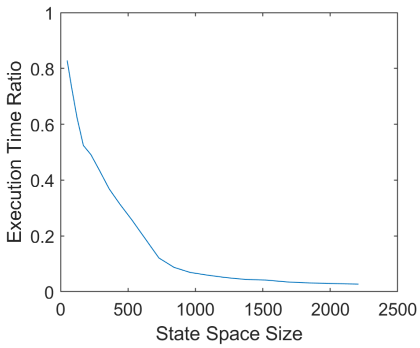

Figure 1: The ratio of the execution time of our improved version over the original algorithm in [20], $K_a = K_d$.

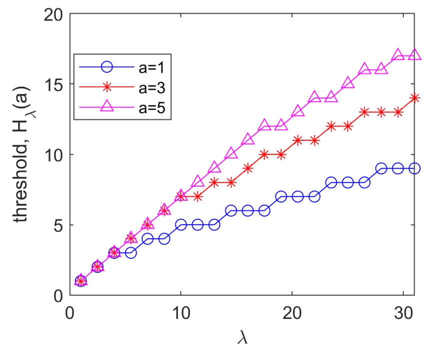

Figure 2: Threshold $H_\lambda(a)$ versus λ for the per-UE stochastic arrival problem at different a. $p_s = 0.5$, $p_g = 0.5$.

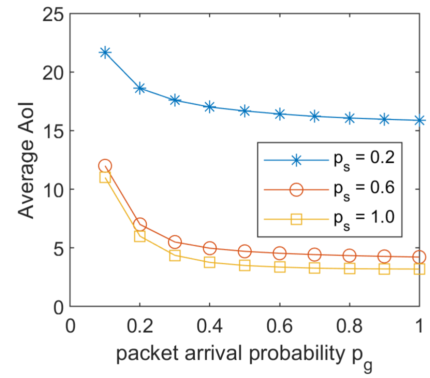

Figure 3: AoI performance of the Whittle index policy on the stochastic arrival under different p_s and p_g.

Now we are ready to verify condition 2 of Theorem 3.

Theorem 11. $T^{\pi}_{a,d}(1) \geq T^{\pi}_{a,d}(0)$, $\forall\ \pi \in \Pi'$ *and* $(a,d) \in \mathcal{S}$.

Proof. Let $\tau_u(0) = -1$. By (13), we have

$$T^{\pi}_{a,d}(u) = \sum_{j=1}^{\infty} \mathbb{E}\Big[\sum_{t=\tau_u(j-1)+1}^{\tau_u(j)} \beta^t u^{\pi}_t \Big| (a_0, d_0) = (a,d), u_0 = u \Big],$$

We now focus on a given j, and compare the weights of the active actions during the intervals leading to the j-th successful transmission for the two chains, i.e., the active actions in $\big[\tau_0(j-1)+1, \tau_0(j)\big]$ for chain 0 and that in $\big[\tau_1(j-1)+1, \tau_1(j)\big]$ for chain 1. We divide the active actions of chain 0 in $\big[\tau_0(j-1)+1, \tau_0(j)\big]$ into 2 sets. Recall from Lemma 10 that $\tau_1(j) \leq \tau_0(j)$. Set $A(j,0)$ contains the active actions of chain 0 during the interval $\big[\tau_0(j-1)+1, \tau_1(j)\big]$ and set $B(j,0)$ contains the rest of the active actions of chain 0 not in this interval. Note that if $\tau_0(j-1)+1 > \tau_1(j)$, then $A(j,0) = \emptyset$. We now do a similar split for the active actions of chain 1 in the interval $\big[\tau_1(j-1)+1, \tau_1(j)\big]$. Recall we have shown in the proof of Lemma 10 that, for each active action in $A(j,0)$, there exists an active action in chain 1 that happens at the same time (see case 2 of the induction step there). This allow us to construct the set $A(j,1)$, which includes all active actions of chain 1 that have a corresponding active action in $A(j,0)$. In other words, the active actions in $A(j,1)$ and $A(j,0)$ occur at exactly the same set of times, and we must have $|A(j,1)| = |A(j,0)|$. Similarly, define $B(j,1)$ to be the rest of the active actions of chain 1 not in $A(j,1)$. Note that by our channel-event coupling, $|A(j,1)|+|B(j,1)| = |A(j,0)|+|B(j,0)|$, which then implies $|B(j,1)| = |B(j,0)|$.

Let ω be an active action and t_ω be the time when ω happens. We can then write $T^{\pi}_{a,d}(1) - T^{\pi}_{a,d}(0)$ as

$$T^{\pi}_{a,d}(1) - T^{\pi}_{a,d}(0) = \sum_{j=1}^{\infty} \Big\{ \mathbb{E}\big[\sum_{\omega \in A(j,1)} \beta^{t_\omega} - \sum_{\omega \in A(j,0)} \beta^{t_\omega} \big] + \mathbb{E}\big[\sum_{\omega \in B(j,1)} \beta^{t_\omega} - \sum_{\omega \in B(j,0)} \beta^{t_\omega} \big] \Big\}. \quad (23)$$

Clearly, the active actions in $A(j,0)$ and $A(j,1)$ have the same weights and are canceled out during the subtraction. The active actions in $B(j,0)$ fall into $\big[\tau_1(j)+1, \tau_0(j)\big]$, while the active actions in $B(j,1)$ fall into $\big[\tau_1(j-1)+1, \tau_1(j)\big]$, which are strictly ahead in time. Thus, we must have (23) ≥ 0 and the theorem then follows. $\square$

Obviously, when $\lambda = +\infty$, no state will choose to activate and when $\lambda = -\infty$, every state will choose to activate. Thus, by Theorem 3 and Theorem 11, we conclude that the per-UE MDP is Whittle indexable. Since our analysis does not depends on K_a or K_d, the indexability holds for arbitrarily truncation thresholds K_a and K_d.

V. Simulation Results

In this section, we will use numerical result to verify our theoretical results and evaluate the performance of Whittle index policy in the stochastic-arrival setting. We first introduce a fast algorithm that we use for computing the Whittle index.

A. Fast algorithm for computing Whittle indices

The algorithm that we use is based on [20, Algorithm 2] and [21], which has $O(K^3)$ time complexity, where K is the number of states. The algorithm adds states iteratively to the passive set. At each iteration, the state with the smallest Whittle index among those not in the passive set is identified, and is added to the passive set. Specifically, at a given iteration, suppose that the current passive set is $\mathcal{P}$. The key idea of [20] is to first compute the so-called Marginal Productivity Index (MPI) for every state $y \in \mathcal{S} \setminus \mathcal{P}$. It then searches and picks the one with the smallest MPI to be the next state added to $\mathcal{P}$. The corresponding MPI value will also be the Whittle index for this newly-added state y. Due to page limits, please refer to [20] for the detailed algorithm.

Similar to how AT condition reduces the policy search space from Π to Π', we improve this algorithm by reducing the search space in each iteration. Specifically, instead of comparing the MPI for every possible $y \in \mathcal{S} \setminus \mathcal{P}$, we only need to focus on a subset of y that can potentially have the minimum MPI. Note that the new passive set $\mathcal{P} \cup y$ must

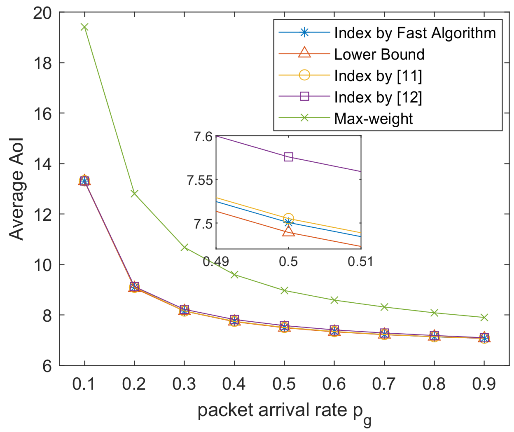

Figure 4: AoI performance versus p_g of Whittle index policy compared to baselines.

also corresponds to an optimal policy [20]. For the stochastic-arrival setting, due to the threshold structure of the optimal policy, if $(a, d) \in \mathcal{P}$ and $(a, d+1) \notin \mathcal{P}$, then we only need to check $y = (a, d+1)$ and we do not need to worry about $(a, d+2), (a, d+3), \ldots$. Even though this improved version still have time complexity of $O(K^3)$, our simulation result (see Figure 1) shows significant reduction in execution time of our improved version compared to the one in [20]. For example, when both a and d are truncated at $K_a = K_d = 30$, the state space is of size 930 and the execution time of our algorithm is only 0.07 of that of [20].

B. Evaluating the Whittle index policy

Next we will present numerical results evaluating the performance of Whittle index policy.

We first verify Lemma 8 and the Whittle indexability of the stochastic-arrival setting by simulation. We use $\beta = 0.99$, $p_s = 0.5, p_g = 0.5, K_a = K_d = 50$ and use value iteration to numerically compute the optimal policy and the passive set. We then verify the threshold property and plot in Figure 2 the threshold $H_\lambda(a)$ as functions of λ for three different values of a. As the figure indicates, for each a, the threshold is monotonically increasing in λ, which confirms both the threshold structure of the optimal policy and the indexability.

We then study the AoI performance in the stochastic-arrival setting under the Whittle index policy. The system is set to have 100 UE and the capacity of the BS is 30. The state space is truncated at 60 for both a and d. We assume each UE has identical p_s and p_g. Note that when $p_g = 1$, the system is equivalent to the generate-at-will setting. In Figure 3, we choose three different values of p_s and plot the total system time-averaged AoI (i.e., $v(h) = h$) as p_g varies. We can observe that when p_g is large, the AoI performance approaches that of $p_g = 1$ (i.e., generate-at-will), but the difference is more prominent when p_g is small.

Finally, we compare the AoI performance of the Whittle index policy with both the lower bound (6) and the performance of a number of baseline polices. The first one (labelled "index in [12]") is an index policy using the closed-form expression in [12]. The second one (labelled "index in [11]") is an index policy using the closed-form expression in [11]. Both expressions provide approximations to the Whittle index. The third one (labelled "Max-weight") is an index policy where the index of each UE n is $p_s(n)\big(a_n(t) + d_n(t)\big)$, i.e., AoI multiplied by the success probability. We note that [11] and [12] use an average-cost MDP formulation, while we use a discounted-cost MDP formulation in this paper. It is well-known that, when β is close to 1 (we use $\beta = 0.99$ in our simulation), the optimal policy for the discounted-cost MDP will approach that for the average-cost MDP. Therefore, below we use the time-averaged AoI as the common metric to compare various policies.

We then simulate the following setting. The number of UE is 100 and the capacity is 30. The channel success probability of each UE is uniformly distributed from the set $\{0.15, 0.25, 0.35, 0.55, 0.85\}$. Note that all UE have the same packet generation rate p_g. Figure 4 shows the total time-averaged system AoI of the various policies as functions of p_g. We can make the following observations. The AoI of our Whittle index policy is close to the lower bound, and is substantially lower than the one using the index of $p_s(n)\big(a_n(t) + d_n(t)\big)$. This confirms the asymptotic optimality of the Whittle index [15] and [16]. Interestingly, the policies using the approximate indices given in [11] and [12] also attain similar AoI as the one using our precise Whittle index. A closer examination reveals that the index policy using our accurately-computed Whittle index achieves the best performance, the index policy using the expression in [11] achieves slightly worse performance and the index policy using the expression in [12] achieves the worst performance. This is interesting because it shows that an imprecise index may also be close to optimal. Thus, one may wish to understand how to quantify the performance guarantees of such policies using imprecise indices, which we leave for future work.

VI. Conclusion

In this paper, we propose the active-time (AT) condition, which is a powerful sufficient condition for Whittle indexability. By a novel coupling approach, we verify the AT condition for the AoI minimization problem under the stochastic-arrival setting, and thus establish its Whittle indexability for the first time in the literature. We also present an improved version of a fast-computing algorithm for computing Whittle indices. Both the AT condition and the coupling approach may be applied to other large RMAB problems. Note that the results in this paper assume a finite state-space. For future work, we will extend our AT condition to infinite state space, and to the partial index formulation in [24] when there are multiple heterogeneous resources.

REFERENCES

[1] M. A. Abd-Elmagid and H. S. Dhillon, "Average peak age-of-information minimization in uav-assisted iot networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 2003–2008, 2019.

[2] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.

[3] Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," in *Proceedings of IEEE INFOCOM*, San Francisco, CA, USA, April 2016, pp. 1–9.

[4] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proceedings of IEEE INFOCOM*, Orlando, FL, USA, March 2012, pp. 2731–2735.

[5] V. Tripathi and E. Modiano, "A Whittle index approach to minimizing functions of age of information," in *57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, USA, Sep. 2019, pp. 1160–1167.

[6] 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.

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

[8] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Can decentralized status update achieve universally near-optimal age-of-information in wireless multiaccess channels?" in *30th International Teletraffic Congress (ITC 30)*, Vienna, Austria, Sep. 2018, pp. 144–152.

[9] B. Sombabu, A. Mate, D. Manjunath, and S. Moharir, "Whittle index for AoI-aware scheduling," in *International Conference on COMmunication Systems and NETworkS (COMSNETS)*, Bengaluru, India, Jan. 2020, pp. 630–633.

[10] J. Sun, Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Closed-form Whittle's index-enabled random access for timely status update," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1538–1551, 2020.

[11] J. Sun, Z. Jiang, S. Zhou, and Z. Niu, "Optimizing information freshness in broadcast network with unreliable links and random arrivals: An approximate index policy," in *The 2nd Workshop on Age of Information (AoI Workshop 2019), in conjunction with IEEE INFOCOM 2019*, April, 2019, pp. 115–120.

[12] Z. Tang, Z. Sun, N. Yang, and X. Zhou, "Whittle index-based scheduling policy for minimizing the cost of age of information," *IEEE Communications Letters*, vol. 26, pp. 54–58, 2022.

[13] M. Chen, K. Wu, and L. Song, "A Whittle index approach to minimizing age of multi-packet information in IoT network," *IEEE Access*, vol. 9, pp. 31 467–31 480, 2021.

[14] A. Maatouk, S. Kriouile, M. Assad, and A. Ephremides, "On the optimality of the Whittle's index policy for minimizing the age of information," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1263–1277, 2021.

[15] R. R. Weber and G. Weiss, "On an index policy for restless bandits," *Journal of Applied Probability*, vol. 27, no. 3, p. 637–648, 1990.

[16] ——, "Addendum to 'on an index policy for restless bandits'," *Advances in Applied Probability*, vol. 23, no. 2, p. 429–430, 1991.

[17] D. Bertsimas and J. Niño-Mora, "Conservation laws, extended polymatroids and multiarmed bandit problems; a unified approach to indexable systems," *Math. Operat. Res. 21*, vol. 21, pp. 257–306, 1996.

[18] J. Niño-Mora, "Restless bandits, partial conservation laws and indexability," *Advances in Applied Probability*, vol. 33, pp. 76–98, March 2001.

[19] K. Glazebrook, D. Hodge, and C. Kirkbride, "Monotone policies and indexability for bidirectional restless bandits," *Advances in Applied Probability*, vol. 45, pp. 51–85, March 2013.

[20] N. Akbarzadeh and A. Mahajan, "Conditions for indexability of restless bandits and an $\mathcal{O}\left(k^3\right)$ algorithm to compute Whittle index," *Advances in Applied Probability*, vol. 54, no. 4, p. 1164–1192, 2022.

[21] ——, "Conditions for indexability of restless bandits and an $\mathcal{O}\left(k^3\right)$ algorithm to compute Whittle index – corrigendum," *Advances in Applied Probability*, p. 1–2, 2023.

[22] P. Whittle, "Restless bandits: activity allocation in a changing world," *Journal of Applied Probability*, vol. 25, no. A, p. 287–298, 1988.

[23] D. R. Jiang and W. B. Powell, "An approximate dynamic programming algorithm for monotone value functions," *Oper. Res.*, vol. 63, no. 6, pp. 1489–1511, 2015.

[24] Y. Zou, K. T. Kim, X. Lin, and M. Chiang, "Minimizing age-of-information in heterogeneous multi-channel systems: A new partial-index approach," in *Proceedings of the Twenty-Second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (ACM MobiHoc)*, New York, NY, USA, 2021, p. 11–20.