Index Policies for RMAB Problems with Varying Capacity and Global State Dependency

Sixiang Zhou*, Xiaojun Lin†, and Lei Jiao‡
* Elmore Family School of Electrical and Computer Engineering, Purdue University,West Lafayette, IN, USA
† Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
‡ Center for Cyber Security and Privacy, University of Oregon, Eugene, OR, USA
Emails: zhou1087@purdue.edu, xjlin@ie.cuhk.edu.hk, ljiao2@uoregon.edu

***Abstract*—Motivated by AoI minimization in wireless networks, we study a variable-capacity version of the restless multi-armed bandit (VC-RMAB) problem, where there is a global state controlling both the available capacity and the agent state transitions. The changing global state arises, e.g., where varying interference levels affect the channel quality globally and/or where the background traffic leads to varying available capacity. Due to the presence of the global state, the classical Whittle index policy cannot be used. We first propose new notions of Variable-Capacity (VC) indexability, VC index and VC index policy. We then introduce a new fluid limit that not only scales the system size but also slows down the state evolution of the agents. Based on this new fluid limit, we establish the asymptotic optimality of the VC index policy. Further, we generalize the Active-Time (AT) condition (a sufficient condition for Whittle indexability) to VC indexability, and we use this new AT condition to establish the VC indexability of the variable-capacity unreliable-channel AoI minimization problem. Simulation results show that our new VC index policy not only approaches asymptotic optimality under our new fluid-limit scaling, but also outperforms other baselines (including an approximate Whittle index policy) in non-asymptotic regimes.**

I. Introduction

Age of Information (AoI) is an important metric capturing the freshness of information, which is crucial for many emerging applications (e.g., Internet of Things (IoT) [1] and Intelligent Transportation Systems [2]). When a base-station (BS) collects information updates from multiple sources through a wireless channel, the resulting AoI minimization problem is usually modeled as a Markov Decision Process (MDP), or more specifically a Restless Multi-Arm Bandit (RMAB) problem [3]–[8]. When the number of sources is large, this problem is known to suffer from the curse-of-dimensionality, i.e., its complexity increases exponentially with the problem size. Whittle index policy [9] is very useful in decomposing such an RMAB problem into sub-problems, which leads to low-complexity and asymptotically optimal solutions. Due to this reason, it has been widely used in the literature of AoI [3]–[8].

This work was supported in part by NSF Awards CNS-2047719, CNS-2113893, CNS-2225949 and CNS-2225950. The research work described in this paper was partly conducted in the JC STEM Lab of Advanced Wireless Networks for Mission-Critical Automation and Intelligence funded by The Hong Kong Jockey Club Charities Trust.

However, the use of Whittle index usually requires that the problem follows some common structures. First, the problem must be weakly-coupled [10], i.e., given the action of an agent, its state-evolution is independent of all other agents in the system. Second, the global capacity constraint that the agents share need to be time-invariant. Unfortunately, many practical problems, especially in wireless networks, do not follow these two structural assumptions. For example, when the sources transmit information to a base-station (BS), the interference and noise level at the BS will affect the transmission success probabilities of all sources [11]–[14]. As a result, the state-evolution of a source given its action is no longer independent of that of other sources. Instead, they are correlated through this common dependency on the global interference/noise level. Further, if the BS also serves other cross-traffic [15]–[17], then the capacity available to the sources is no longer time-invariant. In these situations where either or both of the two structural assumptions are violated, one can no longer use the Whittle index.

In this paper, we are precisely interested in such a setting, i.e., a multi-agent scheduling problem with a time-varying global state that affects both the available capacity in the system and the local state-transition probability of each agent. Following the MDP formulation, we assume that this global state also changes in a Markovian manner. We refer to this problem as the Variable-Capacity Restless Multi-Armed Bandit (VC-RMAB) problem. In the examples above, this global state can capture both the change in the interference and noise level at the BS and the available capacity. We note that our VC-RMAB problem is equivalent to the contextual RMAB problem studied in [18], where our global state can be viewed as the context in [18]. Although [18] proposes a heuristic index policy, it does not provide analytical results on the conditions under which the heuristic index policy is asymptotically optimal (see further discussions in Remark 1). In contrast, our goal in this paper is to extend the Whittle index to this more-general setting and study its asymptotic optimality analytically.

Specifically, based on a similar relaxation as in [18], in this paper we first propose concepts of variable-capacity (VC) index, VC indexability and VC index policy, which generalize the corresponding notions from Whittle index. Unfortunately, the asymptotic optimality of the VC index policy is more difficult to establish due to the global state dependency and capacity variation. Note that for the Whittle index policy, its asymptotic-optimality is usually established under a fluid limit that scales both the number of agents and the capacity by a large factor [19]–[21]. In the standard RMAB problem, such a scaling averages out the stochasticity of each local agent. However, it does not average out the stochasticity of the global state in our VC-RMAB problem. As a result, we cannot establish the asymptotic optimality of the VC index policy under this scaling (See Sec. IV-A for a concrete example). To overcome such difficulty, we propose a new fluid limit that not only scales up the system size, but also slows down the evolution of the per-agent MDPs. Under this new scaling, we are able to show the asymptotic optimality of the VC index policy. Practically, this new scaling implies that our VC index policy is more likely to be close-to-optimal when the global state evolves faster compared to the local state.

We then apply our new VC index to the variable-capacity unreliable-channel AoI minimization problem. By generalizing the Active-Time condition from [22], [23], we show that the VC indexability holds for the AoI minimization problem, and therefore our VC index policy can be used. Through extensive numerical results, we show that the proposed VC index policy not only approaches asymptotic optimality under our new fluid scaling, but also outperforms a trivial extension of the classical Whittle index policy in non-asymptotic regimes.

II. System Model

We first introduce the AoI minimization problem with a global state and variable capacity. Consider a BS that is responsible for collecting the status update from N information sources. Assume that time is slotted, such that each transmission takes exactly one unit of time. The most notable feature that differentiates our VC-RMAB problem from the well-known RMAB problems is a time-varying global state that affects both the available capacity and the agents' state transitions. Specifically, we assume that the global state at time-slot t is $\theta_t \in \Theta$, which is fixed within each time slot t and known at the beginning of the slot. We further assume that the global state θ_t evolves according to a Markov chain with known state transition probabilities given by

$$\Pr\{\theta_{t+1} = \theta' | \theta_t = \theta\} = p_{\theta\theta'}, \ \forall \theta, \theta' \in \Theta. \tag{1}$$

We collect $p_{\theta\theta'}$ into a global state transition matrix $\mathbf{P}^\Theta \in \mathbb{R}^{|\Theta|\times|\Theta|}$. The global state θ_t affects the available capacity through the mapping $C : \Theta \to \mathbb{N}_+$, i.e., given the global state at θ_t, the BS can schedule at most C_{θ_t} sources for transmission time slot t. We further assume that $C_\theta \le N$ for all $\theta \in \Theta$.

Next, we model the status updates in a generate-at-will unreliable-channel AoI minimization problem [3] and describe how the global state also affects the state transitions. Specifically, in the generate-at-will setting, every source always has an up-to-date status-packet available. Denote $u_t^n = 1$ when the BS schedules a transmission from UE n at time t, and $u_t^n = 0$, otherwise. The hard capacity constraint requires that $\sum_n u_t^n \le C_{\theta_t}$ for all time t. However, due to wireless imperfection, each transmission is unreliable, whose success probability also depends on the global state. Specifically, we assume that every packet transmission from source n succeeds with probability $q_\theta^n \in (0, 1]$ if the global state is θ. Define h_t^n to be the AoI of source n at time t, which is the time elapsed from the generation time of the last received packet at the BS to the current time t. In the general-at-will model, the AoI of source n will reduce to 1 if the transmission from source n succeeds, or increase by 1, otherwise. For technical reasons, we restrict the maximum AoI to be $\bar{h}$. Let $\mathrm{nx}(h) = \min\{h+1, \bar{h}\}$. The AoI evolution of source n is thus

$$h_{t+1}^n = \begin{cases} 1 & u_t^n = 1 \text{ and transmission succeeds,} \\ \mathrm{nx}(h) & \text{otherwise.} \end{cases} \tag{2}$$

We can now formulate the AoI minimization problem for the above system as a multi-agent MDP problem, where each agent denotes a source. Recall that the global state θ_t follows a Markov chain that is independent of the scheduling decisions. Let $\mathbf{h}_t = [h_t^1, ..., h_t^N]$ be the vector of all sources' AoI and let $\bar{S}_t = (\mathbf{h}_t, \theta_t)$ be the overall system state at time t. Denote the overall system action as $\bar{U}(t) = \{u_t^1, ..., u_t^N\} \in \bar{\mathcal{A}} \triangleq \{0, 1\}^N$. Let policy $\bar{\pi}$ be a mapping from the system state $\bar{S}$ to the system action $\bar{U}$. Let $v^n(\cdot) : \mathbb{N} \to \mathbb{R}$ be a strictly increasing function representing the penalty cost on the AoI of source n. Let $\mathcal{N} = \{1, ..., N\}$ denote the set of all N sources. Our objective is to minimize the expected average cost of all sources, i.e.,

$$\min_{\bar{\pi}} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{n\in\mathcal{N}} \mathbb{E}^{\bar{\pi}} \left[v^n(h_t^n) \right] \tag{3a}$$

$$\text{s.t.} \ \sum_{n=1}^{N} u_t^n \le C_{\theta_t}, \text{for all time } t, \tag{3b}$$

where $\mathbb{E}^{\bar{\pi}}[\cdot]$ denotes the expectation under policy $\bar{\pi}$.

At this point, it should also be clear that our AoI minimization problem is a special case of a more general RMAB problem with variable capacity, where we can replace the AoI state h_t^n by a general local state s_t^n, replace the AoI penalty in (3) by a general per-stage cost $c^n(s_t^n, \theta_t, u_t^n)$, and replace (2) by arbitrary state transition laws P^0 and P^1. We then refer to the problem below as the Variable-Capacity RMAB (VC-RMAB) problem:

$$\min_{\bar{\pi}} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{\infty} \sum_{n\in\mathcal{N}} \mathbb{E}^{\bar{\pi}} \left[c^n(s_t^n, \theta_t, u_t^n)) \right] \tag{4a}$$

$$\text{s.t.} \ \sum\nolimits_{n=1}^{N} u_t^n \le C_{\theta_t}, \text{for all time } t. \tag{4b}$$

III. Variable-Capacity (VC) Indexability and VC Index Policy

The first step of developing the classical Whittle index usually involves relaxing the (hard) instantaneous capacity constraint (like (4b)) into a (soft) constraint averaged across time. Similar to [18], we relax the capacity constraint (4b) only across the time-slots corresponding to each value of the global state θ. Specifically, we first define $\mathcal{T}_\theta$ as the subset of time slots where $\theta_t = \theta$, i.e., $\mathcal{T}_\theta = \{t = 1, 2, ..., T | \theta_t = \theta\}$. We then relax (4b) across each $\mathcal{T}_\theta$, $\theta \in \Theta$, separately. The resulting relaxed problem becomes

$$\min_{\bar{\pi}} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{n\in\mathcal{N}} \mathbb{E}^{\bar{\pi}} \left[c^n(s_t^n, \theta_t, u_t^n)) \right], \tag{5a}$$

$$\text{s.t. } \limsup_{T\to\infty} \frac{1}{|\mathcal{T}_\theta|} \sum_{t\in\mathcal{T}_\theta} \sum_{n\in\mathcal{N}} \mathbb{E}^{\bar{\pi}} \left[u_t^n \right] \le C_\theta, \forall \theta \in \Theta. \tag{5b}$$

Our new index is then developed based on this new relaxation. Associating dual cost λ'_θ to each (5b) and letting $\lambda_\theta = \lambda'_\theta T / |\mathcal{T}_\theta|$, we obtain the Lagrangian as

$$\min_{\bar{\pi}} \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{n\in\mathcal{N}} \mathbb{E}^{\bar{\pi}} \left[c^n(s_t^n, \theta_t, u_t^n) + u_t^n \lambda_{\theta_t} \right]. \tag{6}$$

Then, we can decompose (6) into per-agent MDPs:

$$\min_{\bar{\pi}} \limsup_{T\to\infty} \frac{1}{T} \sum\nolimits_{t=0}^{T} \mathbb{E}^{\bar{\pi}} \left[c^n(s_t^n, \theta_t, u_t^n) + u_t^n \lambda_{\theta_t} \right]. \tag{7}$$

The dual cost λ_θ can be interpreted as the price of scheduling an agent when the system state is θ. The per-gent problem (7) is also an MDP, but with a much smaller size. We note that (7) differs from the per-agent MDP in classical RMAB problem, as the second term $u_t^n \lambda_{\theta_t}$ now also depends on the time-varying global state θ_t. Interestingly, (7) resembles the partial index [24], a recent generalization of Whittle index to the multiple resource constraints. Specifically, in [24], there are multiple channel-types, each channel-type corresponds to a dual price. In our paper, there are multiple global states, each state corresponds to a dual price. However, unlike [24] where the channel-types are deterministic, here the global state evolves according to a random Markov chain. Due to this reason, the performance analysis later in Section IV will be different.

A. The VC Index Approach

We are now ready to formulate the new notions of VC indexability and VC index policy based on the per-agent MDP (7). Let $\mathcal{S}^n$ be the space of the local state of agent n. Let $\Pi^{*n}_{\vec{\lambda}}$ be the set of optimal policies that attain the minimum in (7) for agent n. Given agent n, global state θ and dual costs $\vec{\lambda}$, define the VC passive set $\mathcal{P}^n_\theta(\vec{\lambda})$ for agent n with respect to θ as follows.

$$\mathcal{P}^n_\theta(\vec{\lambda}) = \left\{ s^n \in \mathcal{S}^n \middle| \exists \pi^* \in \Pi^{*n}_{\vec{\lambda}} \text{ s.t. } \pi^*(s^n, \theta) = 0 \right\}. \tag{8}$$

That is, the VC passive set is the set of local state s^n such that action 0 is an optimal action at state (s^n, θ).

Next we will change the value of λ_θ to $\hat{\lambda}_\theta$ while keeping other dual costs fixed. For ease of notation, we denote $\vec{\lambda} = [\lambda_\theta, \vec{\lambda}_{-\theta}]$, where $\vec{\lambda}_{-\theta}$ denotes all other dual costs except λ_θ. We can then define the VC indexability as follows.

Definition 1 (VC Indexability)**.** *The VC Indexability holds if $\mathcal{P}^n_\theta(\vec{\lambda})$ is monotonically expanding in λ_θ, for all $n = 1, ..., N$, $\theta \in \Theta$ and $\vec{\lambda}_{-\theta} \in \mathbb{R}^{|\Theta|-1}$.*

Assuming that VC indexability holds, we can then define the VC index.

Definition 2 (VC Index)**.** *For agent n, state $s^n \in \mathcal{S}^n$ and dual cost vector $\vec{\lambda}_{-\theta}$ (i.e., dual costs other than λ_θ), the VC index $I^n_\theta(s^n, \vec{\lambda}_{-\theta})$ is defined as*

$$I^n_\theta(s^n, \vec{\lambda}_{-\theta}) = \inf\{\hat{\lambda}_\theta \in \mathbb{R} | s^n \in \mathcal{P}^n_\theta([\hat{\lambda}_\theta, \vec{\lambda}_{-\theta}])\}. \tag{9}$$

Inspired by the Whittle index policy, we aim to schedule users with higher VC indices while satisfying the resource constraint at each time. Note that at time t, the global state θ_t is revealed. Thus, we can formulate our scheduling decision as an optimization problem as follows. Let $z^n_{\theta_t}$ be the binary scheduling decision of agent n and $\omega^n_{\theta_t} = I^n_{\theta_t}(s^n_t, \vec{\lambda}_{-\theta_t})$ be a weight equal to the VC index defined in (9) given the current global state θ_t and local agent state s^n_t. We then solve the following optimization problem, which activates the C_{θ_t} agents with the highest VC indices at time t:

$$\max_{\vec{z}(t)=[z^n_{\theta_t}]} \quad \sum\nolimits_{n=1}^{N} \omega^n_{\theta_t} z^n_{\theta_t} \tag{10a}$$

$$\text{s.t.} \quad \sum\nolimits_{n=1}^{N} z^n_{\theta_t} \le C_{\theta_t}, \tag{10b}$$

$$z^n_{\theta_t} \in \{0, 1\}, \text{ for } n \in \mathcal{N}. \tag{10c}$$

Let $\vec{\nu}(t)$ denote the optimal Lagrange multiplier associated with (10b). Our VC Index policy is presented in Algorithm 1.

Remark 1. We note that our VC index policy in Algorithm 1 is quite different from the heuristic index policy proposed in [18]. The index policy in [18] uses the difference between the state-action value function of the per-agent MDP (7) under the active action and that under the passive action, when the dual costs $\vec{\lambda}$ are the optimal dual costs for the relaxed problem (5). As a result, the index policy in [18] requires knowing the optimal dual costs beforehand. It incurs additional overhead to solve the relaxed problem in order to obtain these optimal dual costs. In contrast, our VC index policy does not require knowing the optimal dual costs. Due to this difference, our VC index requires the crucial condition of VC indexability, which we will study in Section V. Finally, we note that [18] does not provide results on the conditions under which the heuristic index policy is asymptotically optimal. (The analysis in [18] only shows that the relative gap between the relaxed problem and the optimal policy diminishes asymptotically to zero under certain settings.) In contrast, below we will provide a new fluid limit for the asymptotic optimality of our VC index policy.

IV. Performance Analysis

In the literature, the asymptotic optimality of index policies is usually shown through a fluid-limit scaling that increase the system size, i.e., both the number of agents and the capacities, by a large integer Ξ [19]–[21], [24]. As $\Xi \to \infty$, the randomness of each agent's state evolution is averaged out. Then, one can describe the system dynamics under the index policy by a deterministic differential equation, show that its fixed point equals to the solution of the relaxed

Algorithm 1 VC Index policy

1: Initialize $\vec{\lambda}(0)$.
2: **for** each epoch e **do**
3: **for** each iteration t **do**
4: Obtain θ_t and $\vec{s}(t) = [s_t^n]$ and compute $\omega_\theta^n = I_{\theta_t}^n(s_t^n, \vec{\lambda}_{-\theta})$ according to (9) for each agent n.
5: Solve (10) and obtain the scheduling decision $\vec{z}(t)$.
6: Assign resources according to $\vec{u}(t) = [u_t^n] = \vec{z}(t)$.
7: **end for**
8: Update λ_θ for all $\theta \in \Theta$ towards the optimal Lagrange multiplier $\vec{\nu}$ associated with (10b), with step size γ, i.e., $\lambda_\theta(t+1) = \gamma\lambda_\theta(t) + (1-\gamma)\nu_\theta(t)$.
9: **end for**

problem, and thus establish the asymptotic optimality of the index policy. However, as we illustrate below, this standard scaling no longer works for our VC-RMAB problem because it cannot average out the randomness in the global state. We will then propose a new fluid-limit scaling, based on which the asymptotic optimality of the VC index policy can be established.

A. Performance Gap under the Standard Fluid-Limit Scaling

We first use a concrete example to illustrate that, since the standard fluid-limit scaling cannot average out the stochasticity of the global state, there will likely exist a gap between the performance of Algorithm 1 and the solution to the relaxed problem (5). Consider the variable-capacity AoI minimization problem introduced in Section II with space $\Theta = \{0, 1\}$ of the global state. When $\theta_t = 0$, assume that the capacity $C_0 = 0$. When $\theta = 1$, assume that the capacity $C_1 = N/10$, i.e., 10% of sources can be scheduled at a time. For simplicity, suppose that the channel is always reliable, i.e., $q_\theta^n = 1$ for all n and all $\theta \in \Theta$. We further assume that the global state follows a two-state Markov chain that is slowly-varying, e.g., with transition probabilities $p_{00} = p_{11} = 0.999$ and $p_{01} = p_{10} = 0.001$.

Now consider at time τ, we have a global state transition from $\theta_{\tau-1} = 0$ to $\theta_\tau = 1$. Due to the slow-varying nature of our global state, we expect a long period (on average 1000 time slots) of $C_{\theta_t} = 0$ before $\tau - 1$. Hence, we can expect that the AoI of all sources will be very high, e.g., around 1000. Similarly, we can expect a long period of $C_{\theta_t} = N/10$ after τ. For Algorithm 1, it has a hard capacity constraint per-time-slot. Thus, it takes 10 time slots after τ to schedule all sources. At the end of the 10-th time-slot, the AoI of the sources will be between 1 to 10. However, for the relaxed problem (5), it capacity constraint for all time slot with $\theta = 1$ is relaxed. Thus, the solution of (5) may schedule all sources at time τ to transmit. As a result, all sources' AoI may reduce to 1 immediately after τ. Clearly, there is a gap between the decisions under the VC index policy and the solutions to the relaxed problem (5). Note that since the relaxed problem (5) only provides a lower bound for the system cost, the above gap may not directly imply that the VC index policy is not asymptotically optimal. Nonetheless, since we have to rely on the relaxed problem (5) for our performance analysis, it appears difficult to establish the asymptotic optimality of the VC index policy through the above fluid-limit scaling.

B. A New Fluid-Limit Scaling

To address this difficulty, below we propose a new fluid-limit scaling for the VC-RMAB problem, which not only increases the system-size, but also slows down the state-transitions of the agents.

System-size scaling: This is the same as the standard scaling, which increases both the number of agents and the capacity by a large factor $\Xi \in \mathbb{N}^+$. With a slight abuse of notation, in this subsection we will use the original agent index n to denote an agent-type, where each agent-type n now has Ξ instances of agents.

Agent-time scaling: Further, in our new scaling, the local state transition matrix is also being scaled Φ-times slower. Specifically, let $\mathbf{P}_{u\theta} \in \mathbb{R}^{|\mathcal{S}^n|N \times |\mathcal{S}^n|N}$ denote the original state-transition matrix for all agent types under action u and global state θ. In our Φ-scaled system, this transition matrix becomes

$$\mathbf{P}_{u\theta}^{\Phi} = (1 - \frac{1}{\Phi})\mathbf{I} + \frac{1}{\Phi}\mathbf{P}_{u\theta}, \tag{11}$$

where $\mathbf{I}$ is the identity matrix. Note that the transition matrix P^Θ of the global state θ remains unchanged. To compensate for this slow-down of agents' state-transitions, we also let $\tau = t/\Phi$ and replace γ in line 8 of Algorithm 1 by γ/Φ. Due to the above slow-down, however, the optimal dual costs $\vec{\nu}$ for (10b) now also depends on Φ, which we denote by $\vec{\nu}^\Phi$ for the system with the scaling factor Φ.

Fluid limit of Algorithm 1: Below, we will study the fluid-limit of Algorithm 1 by first letting $\Xi \to \infty$ and then letting $\Phi \to \infty$. Letting $\Xi \to \infty$ allows us to analyze the evolution of Algorithm 1 without worrying about the randomness of each agent. Then, letting $\Phi \to \infty$ enables us to derive a deterministic differential equation of the system dynamics under Algorithm 1 without worrying about the randomness in the global state θ_t. This differential equation then reveals the structure of the fixed point of Algorithm 1 at the fluid-limit.

We first derive an equation describing the dynamics of Algorithm 1 when $\Xi \to \infty$. At each time slot t, we define $\vec{\sigma}^\Phi(t) \in \mathbb{R}^{|\mathcal{S}^n|N}$ to be the vector of agent-state occupancy measures, where $\sigma_s^{n\Phi}(t)$ is the fraction of agents of type n that are in state s. Since $\Xi \to \infty$, there are infinitely many agents and the mean field theory can be applied to $\vec{\sigma}^\Phi(t)$. Note that Line 5 of Algorithm 1 can be viewed as a per-time-slot function that maps the current agent-state occupancy measure $\vec{\sigma}^\Phi(t)$, current global state θ_t and current dual costs $\vec{\lambda}$ to a decision vector $\vec{z}(t) \in \mathbb{R}^{\{0,1\} \times |\mathcal{S}^n|N}$, where $z_{su}^n(t)$ denotes the portion of agents in type n and state s that are scheduled to action u. In other words, it can be described as a function f^{vc} such that

$$\vec{z}(t) = f^{\text{vc}}\big(\vec{\sigma}^\Phi(t), \theta_t, \vec{\lambda}(t)\big). \tag{12}$$

Let $\vec{\rho}^{\Phi}_{\theta_t}(t)$ be the vector of agent-state and action measures, i.e., $\rho^{n\Phi}_{s\theta_t u}(t) = z^n_{su}(t)\sigma^{n\Phi}_s(t)$. Then, the dynamics of Algorithm 1 in the limit $\Xi \to \infty$ is

$$\sigma^{n\Phi}_s(t+1) = \sum\nolimits_{d\in\mathcal{S}^n}\sum\nolimits_{u=0}^{1}\rho^{n\Phi}_{d\theta_t u}(t)p^{n\theta_t u\Phi}_{d\to s}, \forall n, s \quad (13a)$$

$$\theta_t \text{ evolves according to } \mathbf{P}^{\Theta} \text{ in (1)}, \quad (13b)$$

$$\lambda_\theta(t+1) = -\gamma_t\big(\lambda_\theta(t) - \nu^{\Phi}_\theta\big), \quad (13c)$$

$$\vec{\nu}^{\Phi} \text{ is the optimal dual vector associated with (10b)}, \quad (13d)$$

where $p^{n\theta u\Phi}_{d\to s}$ is the local state-transition probability from d to s, for agent type n under global state θ, action u and scaling factor Φ. Note that given θ_t, (13a) is deterministic even when Φ is finite because of $\Xi \to \infty$. However, for finite Φ, the stochasticity of θ_t still makes $\vec{z}(t)$ random (see (12)). Hence, the evolution of $\vec{\sigma}^{\Phi}(t)$ is still random and cannot be described by a deterministic differential equation.

Fortunately, such randomness can be averaged out by the agent-time scaling as we let $\Phi \to \infty$. Recall that our gent-time scaling both slows down the agent-state evolution $\vec{\sigma}(t)$ by Φ and shrinks time by $\tau = t/\Phi$. When $\Phi = 1$, define $\vec{F}^{\text{vc}}$ to be

$$\vec{F}^{\text{vc}}(\vec{\sigma}, \theta, \vec{\lambda}) \triangleq \Big[\sum\nolimits_{d\in\mathcal{S}^n}\sum\nolimits_{u=0}^{1}\rho^{n1}_{d\theta_u}p^{n\theta u}_{d\to s}\Big]_{n,s} - \vec{\sigma}. \quad (14)$$

In other words, $\vec{F}^{\text{vc}}(\vec{\sigma}, \theta, \vec{\lambda})$ is the right-hand-side of (13a) minus the current $\vec{\sigma}$. To get rid of the dependency on θ_t, note that as Φ increases, each small time-interval $\Delta\tau$ (after shrinking) corresponds to a long interval $\Phi\Delta\tau$ in the original time. Thus, as $\Phi \to \infty$, the random global states seen in the time-interval $[\tau, \tau + \Delta\tau]$ will span the stationary distribution of θ according to P^{Θ}. Using this idea, we can then show that, as $\Phi \to \infty$, the limiting dynamics of Algorithm 1 satisfies

$$d\vec{\sigma}^{\infty}/d\tau = \mathbb{E}_\theta\Big[\vec{F}^{\text{vc}}\big(\vec{\sigma}^{\infty}, \theta, \vec{\lambda}^{\infty}\big)\Big], \quad (15)$$

where the expectation is taken with respect to the stationary distribution of the global state θ. For detailed steps, please refer to our technical report [25]. Hence, a fixed point $\vec{\sigma}^*$ for the fluid-limit of Algorithm 1 must then satisfy

$$\mathbb{E}_\theta\Big[\vec{F}^{\text{vc}}\big(\vec{\sigma}^*, \theta, \vec{\lambda}^*\big)\Big] = 0. \quad (16)$$

Further, from (13c), such a fixed point should satisfy $\vec{\lambda}^* = \vec{\nu}^*$ with $\vec{\nu}^*$ satisfies (13d).

Fluid limit of the relaxed problem: We next move to the fluid-limit of the relaxed problem (5). Since the relaxed problem (5) can be decomposed into the per-agent MDPs (7), its solution is independent of Ξ. Thus, we only need to consider the limit as $\Phi \to \infty$. Similar to the fluid-limit analysis of Algorithm 1, when $\Phi \to \infty$, the global state will evolve much faster than the agent's local state. Therefore, when we write down the balance equation of each agent-type, we can assume that the agent sees the stationary distribution of the global state θ. Let $c^n_{s\theta u}$ be the per-stage cost at state (s, θ) of agent n under action u. Define μ^n_s as the fraction of agents n that are in type n and at state s in the fluid limit and denote $x^n_{s\theta u} \in [0, 1]$ as the fraction of agents of type n that takes action u according to the fluid-limit solution of (5). Define α_θ to be the stationary distribution of the global state θ. Then, we can show that, as $\Phi \to \infty$, the fluid limit of (5) can be written as the following optimization problem (detailed derivations are available in [25]):

$$\min_{\vec{x},\vec{\mu}} \quad \sum_{n\in\mathcal{N}}\sum_{s\in\mathcal{S}^n}\sum_{\theta\in\Theta}\sum_{u=0}^{1} c^n_{s\theta u}x^n_{s\theta u}\mu^n_s\alpha_\theta \quad (17a)$$

$$\text{s.t.} \quad \sum\nolimits_{n\in\mathcal{N}}\sum\nolimits_{s\in\mathcal{S}^n} x^n_{s\theta 1}\mu^n_s \le C_\theta/N, \forall\theta\in\Theta, \quad (17b)$$

$$x^n_{s\theta u} \ge 0 \text{ and } \sum\nolimits_{u=0}^{1} x^n_{s\theta u} = 1, \forall n, s, \theta, u, \quad (17c)$$

$$\sum\nolimits_{n\in\mathcal{N}}\sum\nolimits_{s\in\mathcal{S}^n}\mu^n_s = 1 \quad (17d)$$

$$\mu^n_s = \sum_{\theta\in\Theta}\alpha_\theta\sum\nolimits_{u=0}^{1}\sum_{d\in\mathcal{S}^n} x^n_{d\theta u}\mu^n_d p^{nu\theta}_{d\to s}, \forall n, s, \quad (17e)$$

Note that by letting $X^n_{s\theta u} = x^n_{s\theta u}\mu^n_s\alpha_\theta$ (which can be viewed as the agent-state-and-action occupancy measure in the fluid limit) and by replacing (17c) by $\sum_{u=0}^{1} X^n_{s\theta u} = \mu^n_s$, (17) can be easily converted into an LP.

Asymptotic optimality: Next, we wish to show that the fixed point of Algorithm 1 and the solution to the relaxed problem (5) coincide in our new fluid limit. Here we need one additional assumption on the VC index. Denote $I^{n\Phi}_\theta(s, \vec{\lambda}_{-\theta})$ as the VC index for the scaled relaxed problem with $\Xi \to \infty$ and Φ finite, which can be computed by (9) based on a version of the relaxed problem (5) scaled by a finite Φ. Similarly, at $\Phi = +\infty$, by decomposing (17) through Lagrange duality, we can also derive a VC index for $\Phi = +\infty$. Below, we make an assumption that relates the VC index for finite Φ and the VC index for $\Phi = +\infty$ (i.e., based on (17)), which will be verified numerically later.

Assumption 3. *$I^{n\infty}_\theta(s, \vec{\lambda}_{-\theta}) = \lim_{\Phi\to\infty} I^{n\Phi}_\theta(s, \vec{\lambda}_{-\theta})$ exists and equals to the VC index based on* (17)*, for all $n \in \mathcal{N}$, $s \in \mathcal{S}^n$ and $\theta \in \Theta$.*

With Assumption 3, we can then compare the fixed point of Algorithm 1 with the solution of (17) at the fluid limit. Specifically, from (13c), (13d) and (16), we can summarize properties of the fixed point of Algorithm 1 at the fluid limit. Then, we can show that, under Assumption 3 and by the VC indexability for all $\Phi \in \mathbb{N}_+$, those properties imply the KKT conditions of (17). We then obtain the following lemma.

Lemma 4. *Suppose that both Assumption 3 and the VC indexability as defined in Definition 1 hold. Then, any fixed point of Algorithm 1 at the fluid-limit corresponds to a solution of* (17).

For the complete proof of Lemma 4, see [25]. From here, it is then not hard to show that the relative performance gap between our proposed VC index policy and the lower bound (17) approaches zero at our fluid limit, under a global attractor assumption similar to that of Whittle index [9]. See [25] for detailed derivations.

V. AoI Minimization under the Variable Capacity Setting

In this section, we return to the AoI minimization problem (3). We will show that VC indexability holds for this problem at every Φ, and as a result the VC index policy can be directly used. Recall that a key difference from Whittle index is that the global state θ_t also changes in the per-agent MDP (7). As a result, verifying VC indexability is more involved than verifying Whittle indexability. Below, we first extend an Active Time (AT) condition for Whittle indexability [22], [23] to VC indexability, based on which we will establish the VC indexability of our AoI problem.

Towards this end, in this section we will focus on the discounted-cost version of the per-agent MDP (7), with a discount factor $\beta \in (0, 1)$. The AT condition is easier to define under such a discounted-cost setting. Further, we can show that if the AT condition for the discounted-cost version holds for all β close to 1, the VC indexability for the average-cost MDP (7) also holds [25].

We first generalize the AT condition to the VC indexability. Define $Q^{\pi}_{\vec{\lambda}}(s, \theta, u)$ to be the state-action value function for the discounted-cost version of (7) under dual costs $\vec{\lambda}$ and policy π, from initial state (s, θ) and initial action u. Following [22], [23], we have $Q^{\pi}_{\vec{\lambda}}(s, \theta, u)$ being a linear function on $\vec{\lambda}$, i.e.,

$$Q^{\pi}_{\vec{\lambda}}(s, \theta, u) = C^{\pi}_{s\theta}(u) + \sum\nolimits_{\vartheta \in \Theta} T^{\pi}_{s\theta}(u, \vartheta)\lambda_{\vartheta}, \quad (18)$$

where $C^{\pi}_{s\theta}(u)$ does not depend on $\vec{\lambda}$, only the second term depends on $\vec{\lambda}$, and we refer to the factor

$$T^{\pi}_{s\theta}(u, \vartheta) \triangleq \mathbb{E}^{s\theta u}_{\pi}\Big[\sum\nolimits_{t=0}^{\infty} \beta^t \mathbf{1}\{\theta_t = \vartheta\} \cdot \mathbf{1}\{u_t = 1\}\Big] \quad (19)$$

as the ϑ-active time. In other words, $T^{\pi}_{s\theta}(u, \vartheta)$ is the the expected discounted total number of activations when $\theta_t = \vartheta$, given the initial state (s, θ), the initial action u and the policy π. We can now introduce the AT condition for VC indexability. Let Π' be a policy set containing all possible optimal policies.

Definition 5 (The AT condition)**.** *$T^{\pi}_{s\theta}(1, \theta) \geq T^{\pi}_{s\theta}(0, \theta)$ holds for all initial state $s \in \mathcal{S}$, all $\theta \in \Theta$ and all $\pi \in \Pi'$.*

Theorem 6. *Suppose that the AT condition in Definition 5 holds for all β close to 1. Then, the VC indexability holds.*

The proof of Theorem 6 is similar to the proofs in [22], [23] and is available in our technical report [25].

Remark 2. Readers familiar with our partial-index work in [23], [24] may see similarity between the AT condition for the VC index and that for the partial index, in the sense that the global state θ appears to play the role of different resources in [23]. One difference, however, is that in Definition 5, the same θ appears both as the initial state and the state where activations are counted. In contrast, for partial indexability, the AT condition in [23] also requires counting the number of activations of an action different from the initial action. Thus, the AT condition in this paper is structurally different from that in [23].

A. Proving the AT Condition for the AoI minimization Problem

Next, we fix Φ, and show that the AT condition in Def. 5 holds. Given Φ, the state evolution for the AoI of agent n is as follows. Consider $h_t = h$ and $\theta_t = \theta$. If $u_t = 1$, we have

$$h_{t+1} = \begin{cases} h & \text{w.p. } \frac{\Phi-1}{\Phi} \\ \text{nx}(h) & \text{w.p. } \frac{1-q_\theta}{\Phi} \\ 1 & \text{w.p. } \frac{q_\theta}{\Phi}. \end{cases} \quad (20)$$

If $u_t = 0$, AoI will grow to $\text{nx}(h)$ with probability $1/\Phi$. The evolution of the global state θ_t still follows (1). Similar to [22]–[24], we can then show a threshold structure of the optimal policy.

Lemma 7. *For a fixed Φ, all optimal policies π^* of (7) have a threshold structure with respect to h, i.e., for each θ we can find a threshold $H_\theta \in \{1, ..., \bar{h}+1\}$ such that $\pi^*(h, \theta) = 1$ if $h \geq H_\theta$ and $\pi^*(h, \theta) = 0$ if $h < H_\theta$.*

Let Π' be the set of threshold-type policies satisfying Lemma 7. We can then show that the AoI minimization problem satisfies the AT condition, and thus VC indexability holds. For detailed proofs, see [25].

Theorem 8. *Under any given Φ, the variable-capacity AoI minimization problem satisfies the AT condition.*

VI. Simulation Results

We use the following simulation setup for an AoI minimization problem to evaluate our VC index policy. We consider a system with $N = 50$. There are 5 groups of sources, each group has 10 instances. The global state can vary in the set $\Theta = \{1, 2, 3\}$, where the corresponding capacity C_θ is $\{5, 15, 25\}$, respectively. We consider the following state-transition matrix of θ:

$$P^{\Theta} = \begin{bmatrix} 0.8 & 0.05 & 0.15 \\ 0.35 & 0.1 & 0.55 \\ 0.3 & 0.5 & 0.2 \end{bmatrix}. \quad (21)$$

We consider the max AoI of $\bar{h} = 30$. For source groups 1 to 5, their transmission success probabilities under $\theta = 1$ are given by a vector $\vec{p}^1 = [0.1, 0.3, 0.5, 0.7, 0.9]$. For other $\theta \in \Theta$, the transmission success probability vector is obtained by circularly shifting $\vec{p}^1$ to the right by $\theta - 1$ positions. The AoI cost function is simply $v(h) = h$. To verify the asymptotic optimality, we scale the system by positive integers Ξ and Φ as stated earlier.

VC Index Convergence at the Fluid Limit: We start with verifying the convergence of the VC index as $\Phi \to \infty$, i.e., Assumption 3. Figure 1 shows the VC index values for global state $\theta = 1$ and different local states h of the source type 1 as the agent-time scaling factor Φ increases. The dashed line is the value of I^{∞}_{θ} for each state h based on the dual decomposition of the fluid-limit relaxed problem (17). From the figure, we can observe that the VC index I^{Φ}_{θ} converges to the VC index of (17), which verifies Assumption 3.

Equivalence of the Dual Costs: We next verify the convergence of the dual costs, i.e., whether the dual costs of

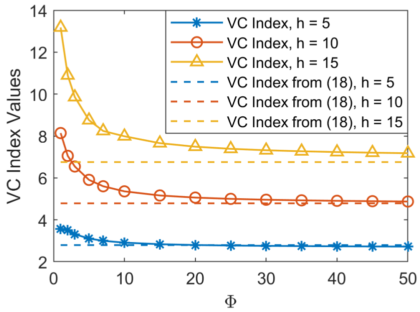

Figure 1: VC index values of different states versus the agent-time scaling factor Φ.

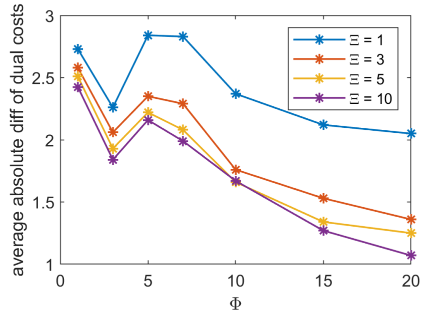

Figure 2: Average absolute difference of the dual costs on different scaling factors Φ and Ξ.

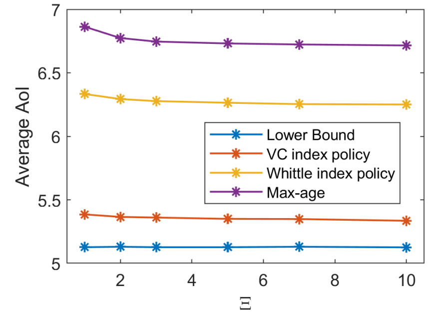

Figure 3: Average AoI of the system under different policies, for $\Phi = 1$ and different Ξ.

Algorithm 1 at the fluid limit equal to the dual costs of the relaxed problem (17) at the fluid limit when $\Xi, \Phi \to \infty$. For Algorithm 1, running it will produce the dual costs. For the relaxed problem (17), we can find the optimal dual costs by a super-gradient descent algorithm on its dual. We then compare the average absolute difference (i.e., L1-norm divided by the vector size) between the final convergent point of the dual costs of our Algorithm 1 and the optimal dual costs of the relaxed problem (17). Figure 2 shows this average absolute difference as Ξ and Φ vary. From the figure, we can observe that we need both large Ξ and Φ for the absolute difference to be small. This confirms that only using the standard system-size scaling (i.e., letting $\Xi \to \infty$) is insufficient; our agent-time scaling is also critical in order for the fixed point of Algorithm 1 to coincide with the solution of the relaxed problem. We note that the average absolute difference between the dual costs is about 1 when $\Phi = 20$, which matches with Figure 1 where the difference between the VC index at $\Phi = 20$ and the limiting index at $\Phi = \infty$ is about that value.

AoI Performance Evaluation Next, we evaluate the average AoI performance of our proposed VC index policy. We will evaluate Algorithm 1 with the following baselines.

- *Max-Age policy.* This policy will greedily activate agents with highest AoI values.
- *(Approximate) Whittle Index policy.* We can compute an approximate Whittle index based on the agent dynamics with the global state averaged out. That is, the transmission success probability of each agent n is $\mathbb{E}_\theta[q_\theta^n]$, with the expectation taken over the steady-state distribution of θ. It turns out that this Whittle index equals to the limiting VC index computed based on (17).
- *Lower Bound.* This baseline comes from the solution of the relaxed problem (5), which serves as the lower bound.

In Figure 3, we fix $\Phi = 1$ and vary Ξ, and compare the average AoI of the system under Algorithm 1 with that of the baselines. From Figure 3, we observe that the average AoI of all polices become closer to the *Lower Bound* as Ξ increases. However, even for large Ξ, there is a noticeable gap between the VC index and the lower bound, which suggests again that only scaling Ξ is insufficient for asymptotic optimality. Further, at $\Phi = 1$ we observe that both the *Max-age* policy and the *Whittle index* policy are worse than the VC index policy by roughly 20% and 10%, respectively, in terms of average AoI, which suggests that the VC index policy is better at dealing with changing capacity and global state dependency.

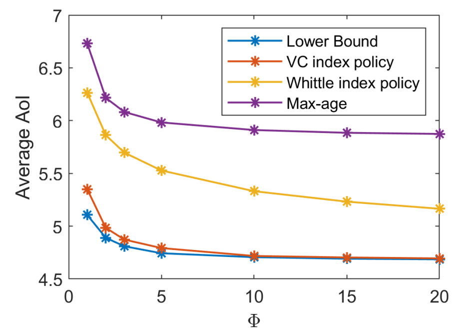

Figure 4: Average AoI of the system under different policies, for $\Xi = 5$ and different Φ.

In Figure 4, we fix $\Xi = 5$ and vary Φ, and also compare the average AoI of Algorithm 1 with that of the baselines. The VC index policy performs consistently better than both the Whittle index policy and the Max-age policy. Further, we observe three points. First, all policies achieve better average AoI as Φ increase, which suggests that the randomness from the global state and variable capacity is harmful to the overall performance. Second, the performance of the VC index policy converges quickly to that of lower bound as Φ increases. This matches with our asymptotic optimality results. Third, even though the Whittle index policy performs closer to the VC index policy as Φ increases, there is a large gap between them. Again, this is because the VC index policy handles the global state variations better for finite Φ.

Readers may question the practical meaning of our new fluid limit with $\Phi \to \infty$, as in practice we cannot slow down the agent dynamics arbitrarily. We conjecture that $\Phi \to \infty$ corresponds to the global state dynamics evolving fast. To evaluate this conjecture, in Fig. 5 we evaluate how P^{Θ} affects the average AoI. Let $\mathbf{I}$ and $\mathbf{1}$ be the 3×3 identity matrix and the matrix of all ones, respectively, and let $\mathbf{P}^{\text{fast}} = 0.5(\mathbf{1} - \mathbf{I})$.

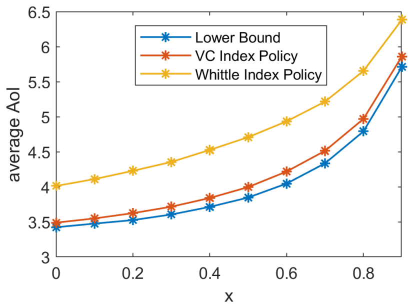

Figure 5: Average AoI of the system under different P^{Θ}.

We then repeat the above simulation while replacing P^{Θ} in (21) by $x\mathbf{I} + (1-x)\mathbf{P}^{\text{fast}}$, for a varying $x \in [0, 1)$. Note that we will keep a fixed $\Phi = 1$. Still, a lower x corresponds to faster global state dynamics. From Figure 5, we observe that the gap between the lower bound and the VC index policy decreases when x approaches 0 (i.e., when the global state evolves faster), which aligns with our conjecture. Interestingly, we observe a similar trend when x approaches 1 (i.e., when the global state evolves very slowly). This may suggest an opposite direction of scaling, which we leaves for future work. Across all x, we observe that the VC index policy consistently outperforms the Whittle index policy, confirming the advantage of our VC index in non-asymptotic regimes.

VII. Conclusion

In this paper we study the VC-RMAB problem where a time-varying global state controls both the available capacity and the agents' state transitions. We propose new notions of the VC indexability and VC index policy. We then introduce a new fluid limit to establish the asymptotic optimality of the VC index policy. We further generalize a sufficient condition for Whittle indexability (the AT condition) to VC indexability, based on which we establish the VC indexability of the variable-capacity unreliable-channel AoI minimization problems. For future work, we plan to establish the asymptotic optimality for another direction of the scaling, i.e., slowing down the global state evolution as suggested by Fig. 5.

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] D. Adelman and A. J. Mersereau, "Relaxations of weakly coupled stochastic dynamic programs," *Oper. Res.*, vol. 56, no. 3, p. 712–727, May 2008. [Online]. Available: https://doi.org/10.1287/opre.1070.0445

[11] X. Li, Q. Zhao, X. Guan, and L. Tong, "Optimal cognitive access of markovian channels under tight collision constraints," in *2010 IEEE International Conference on Communications*, 2010, pp. 1–5.

[12] I.-H. Wang, C. Suh, S. Diggavi, and P. Viswanath, "Bursty interference channel with feedback," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 21–25.

[13] C.-L. Chen, J.-T. Liu, Y.-C. Hsiao, C.-M. Chen, H.-C. Tsao, and C. Lin, "Robust beamforming design for MISO bursty interference channels under traffic uncertainties," *IEEE Transactions on Signal Processing*, vol. 67, no. 6, pp. 1477–1492, 2019.

[14] A. Konrad, B. Y. Zhao, A. D. Joseph, and R. Ludwig, "A Markov-based channel model algorithm for wireless networks," in *Proceedings of the 4th ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, ser. MSWIM '01. New York, NY, USA: Association for Computing Machinery, 2001, p. 28–36. [Online]. Available: https://doi.org/10.1145/381591.381602

[15] Y. Inoue and M. Mandjes, "Characterizing the age of information with multiple coexisting data streams," *IEEE Transactions on Information Theory*, vol. 71, no. 6, pp. 4732–4753, 2025.

[16] J. Xu, I.-H. Hou, and N. Gautam, "Age of information for single buffer systems with vacation server," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1198–1214, 2022.

[17] B. Buyukates and S. Ulukus, "Age of information with Gilbert-Elliot servers and samplers," in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, 2020, pp. 1–6.

[18] X. Chen and I.-H. Hou, "Contextual restless multi-armed bandits with application to demand response decision-making," in *63rd IEEE Conference on Decision and Control (CDC)*, Milan, Italy, 2024, pp. 2652–2657.

[19] 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.

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

[21] I. M. Verloop, "Asymptotically optimal priority policies for indexable and nonindexable restless bandits," *Ann. Appl. Probab.*, vol. 26, no. 4, pp. 1947–1995, 2016.

[22] S. Zhou and X. Lin, "An easier-to-verify sufficient condition for Whittle indexability and application to AoI minimization," in *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications*, 2024, pp. 1741–1750.

[23] ——, "On the active-time condition for partial indexability and application to heterogeneous-channel AoI minimization," in *Proceedings of the Twenty-Fifth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, ser. MobiHoc '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 311–320.

[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)*, Shanghai, China, 2021, p. 11–20.

[25] S. Zhou, X. Lin, and L. Jiao, "Index policies for RMAB problems with varying capacity and global state dependency," Chinese University of Honk Kong, Tech. Rep., 2026. [Online]. Available: https://staff.ie.cuhk.edu.hk/%7exjlin/papers.html