Age-of-Information Aware Scheduling under Markovian Energy Arrivals

Bejjipuram Sombabu and Sharayu Moharir
Department of Electrical Engineering, Indian Institute of Technology Bombay
e-mail: ssombabu@gmail.com, sharayum@ee.iitb.ac.in

***Abstract*—We consider the task of scheduling updates from multiple sources to a central monitoring station via a shared communication channel. Each source harvests energy from nature to measure a time-varying quantity and report these measurements to the monitoring station. Prior work in this area focuses on the setting where energy arrivals are assumed to be independent across time. Motivated by the time-correlation in energy generated by many renewable energy sources, we use a Markov process to model the energy arrivals. The goal is to minimize the time average of the weighted sum of the ages-of-information of the sources. We use Whittle's relaxation and propose a modification of the Whittle Index to design a scheduling policy. We show that our policy outperforms other natural policies via simulations.**

I. Introduction

Age-of-Information (AoI) [1] is a metric that quantifies the freshness of information available at the intended destination. It is defined as the time elapsed since the destination received the latest update from the source. This metric has been used in a variety of applications including scheduling, channel state estimation, caching, and energy harvesting. Refer to [2] for a comprehensive survey of AoI-based works.

In this work, we focus on the task of scheduling updates from multiple sources to a monitoring station via a shared communication channel in order to minimize the time-average of the weighted sum of the AoIs. Each source measures a time-varying quantity. The sources rely on energy harvested from nature to collect measurements and send them to the monitoring station. Each source is equipped with a finite battery to store energy for future use.

AoI-aware scheduling for energy harvesting nodes has been studied in [3]–[9]. These works focus on the setting where energy arrivals are independent across time. However, energy arrivals from renewable sources like the wind are time-correlated [10], [11]. Motivated by this, we focus on the setting where energy arrivals are correlated across time, more specifically, we use a Markov process to model energy arrivals. Such models are known to be good fits for energy arrivals from renewable sources and have been used in [10], [11]. Another closely related work is [12] where channel realizations are Markovian while the sources always have enough energy to transmit.

We use the Whittle's relaxation [13] of the scheduling problem. The key analytical challenge is to show that the problem is indexable and then compute the Whittle Index to design a schedule. The key contributions of our work can be summarized as follows.

978-1-7281-8895-9/20/$31.00 ©

We first show that the scheduling problem is indexable under Markovian energy arrivals. Next, we identify the technical challenges in computing the Whittle Index for our problem. Given these challenges, we propose a modification of the Whittle Index and use it to design a scheduling policy. Via simulations, we show that the proposed policy outperforms natural scheduling policies like greedy scheduling.

II. Setting

A. Network Model

We consider a system with n sources and a monitoring station. Each source relies on energy harvested from nature to collect measurements and send updates to the monitoring station. Each source is equipped with a battery of unit size. Each source consumes one unit of energy to measure and send an update to the monitoring station. Updates are sent via an error-free link. We consider slotted time such that at most one source can send updates to the monitoring station in each time-slot.

B. Energy Arrival Process

Let $\Lambda_i(t) \in \{0,1\}$ be the energy arrival at Source i in time-slot t. We consider two energy arrival models.

Assumption 1: (i.i.d. arrivals) For $1 \leq i \leq n$ and $t \in \mathbb{Z}^+$,

$$\Lambda_i(t) = \begin{cases} 1 & \text{w.p. } p_i \\ 0 & \text{otherwise.} \end{cases}$$

Assumption 2: (Markovian arrivals – Gilbert-Elliot model) For $1 \leq i \leq n$ and $t \in \mathbb{Z}^+$,

$$\Lambda_i(t+1) = \begin{cases} 1 & \text{w.p. } p_i \text{ if } \Lambda_i(t) = 1 \\ 0 & \text{w.p. } (1-p_i) \text{ if } \Lambda_i(t) = 1 \\ 1 & \text{w.p. } (1-q_i) \text{ if } \Lambda_i(t) = 0 \\ 0 & \text{w.p. } q_i \text{ if } \Lambda_i(t) = 0. \end{cases}$$

C. Sequence of Events in a Time-slot

In each time-slot, at most one source is scheduled for communication. This source collects a fresh measurement and sends it to the monitoring station. This is followed by potential energy arrivals to the n sources. This energy is stored in the sources' batteries subject to capacity constraints.

Let $B_i(t)$ denote the energy stored in the battery of Source i at the beginning of time-slot t and $D(t)$ denote the index of the source scheduled for communication in time-slot t. Let $D(t) = 0$ if none of the n sources are scheduled for communication in time-slot t. Since the process of collecting a measurement and sending an update to the monitoring station requires one unit of energy, $B_i(t)$ evolves as follows:

$$B_i(t+1) = \max\{\Lambda_i(t), \big(B_i(t) - \mathbb{1}_{\{D(t)=i\}}\big)^+\}.$$

D. Performance Metric and Goal

We are interested in the age-of-information of the sources at the monitoring station. It is the amount of time elapsed since the monitoring station received the latest update from a source.

Definition 1 (Age-of-Information (AoI)): Let $X_i(t)$ denote the AoI of Source i at the monitoring station at the beginning of time-slot t and $U_i(t) = \max\{\tau : 1 \le \tau \le t - 1 \text{ and } D(\tau) = i\}$ denote the index of the time-slot in which the monitoring station received the latest update from Source i. Then, $X_i(t) = t - U_i(t)$.

In each time slot, a central scheduler determines which source sends an update to the monitoring station. A scheduling algorithm, θ is a time-sequence of such scheduling decisions. Given w_i, i.e., the weight of Source i, for $1 \le i \le n$, the objective is to design a scheduler θ to minimize the time average of the weighted sum of AoIs, i.e.,

$$\limsup_{T\to\infty} \frac{1}{T} \mathrm{E}_\theta \left[\sum_{t=1}^{T} \sum_{i=1}^{n} w_i X_i(t) \right]. \quad (1)$$

Note that E_θ denotes the expected value of the weighted sum of AoIs when the scheduling algorithm θ is employed.

III. Preliminaries

As an intermediate step towards designing a scheduling algorithm to minimize the time-average of the weighted sum of AoIs (1), we construct a new scheduling problem decoupled across the n sources into n sub-problems using Whittle's relaxation of the RMAB problem [13]. In this new construction, each source pays a cost of c units to send an update to the monitoring station. In the rest of this discussion, we drop the source index for convenience.

1) States and Actions: We define the state of a source in time-slot t by $s(t) = (X(t), B(t), \Lambda(t))$. Let the action of the source in time-slot t be denoted by $a(t) \in \{0, 1\}$, where $a(t) = 1$ indicates that the source is scheduled for communication in time-slot t and $a(t) = 0$ means that it remains idle.

2) Cost: The cost incurred by a source in a time-slot, denoted by $C\big(s(t), a(t)\big)$, has two components. One is the cost incurred due to the AoI of the source. The other is a communication cost of c units, which the source pays when it sends an update to the monitoring station. Formally,

$$C\big(s(t), a(t)\big) = w\big(X(t) + 1 - X(t)a(t)B(t)\big) + ca(t).$$

Definition 2: (Cost-Optimal Policy) A policy μ is cost-optimal if it minimizes the average cost defined as follows

$$\limsup_{T\to\infty} \frac{1}{T} \mathbb{E}_\mu \left[\sum_{t=1}^{T} C\big(s(t), a(t)\big) \right]. \quad (2)$$

We consider the sub-problem with one source and the monitoring station and calculate the Whittle index [13] for that source by considering a cost of c for scheduling an update from the source.

Definition 3: (Indexability [13]) Given cost c, let $S(c)$ be the set of states for which the optimal action for the states is to idle. The subproblem is indexable if the set $S(c)$ monotonically increases from the empty set to the entire state space, as c increases from $-\infty$ to ∞.

Definition 4: (Whittle Index [13]) The Whittle index is defined as the cost c that makes both actions for state s equally desirable.

IV. Main Results and Discussion

In this section, we state and discuss our key results. Proofs of our results are discussed in Section V. We first present our results for the setting where energy arrivals are i.i.d. across time-slots (Assumption 1).

A. i.i.d. Energy Arrivals

Our first result focuses on the sub-problem defined in Section III for i.i.d. energy arrivals.

Theorem 5: Under Assumption 1 with $\Lambda(t) \sim \mathrm{Bernoulli}(p)$, the sub-problem defined in Section III is indexable and the Whittle Index for state (x, b, λ) is given by

$$I(x, b, \lambda) = \begin{cases} 0 & \text{if } b = 0; \\ \frac{A}{B} & \text{if } b = 1, \end{cases} \quad (3)$$

where $A = w(((1-p)^x + 1)p^2x^2 + (1 - (1-p)^x)p^2x - 2(1-p)^x)$ and $B = p^2(2 - 2(1-p)^x)$.

We now propose a scheduling policy called Whittle–iid. In each time-slot, this policy schedules the source with the highest Whittle Index.

Algorithm 1: Whittle-iid Scheduling

Input: Number of sources n, (x_i, b_i, λ_i), p_i, $\forall i$.
1 **procedure** In each time-slot, calculate Whittle index $I(x_i, b_i, \lambda_i), \forall i$ using (3).
2 Schedule Source i^* for communication, where $i^* = \arg \max_{i\in\{1,2,\cdots,n\}} I(x_i, b_i, \lambda_i)$.
3 **end procedure**

B. Markovian Energy Arrivals

We now present our results for Markovian energy arrivals (Assumption 2).

Theorem 6: Under Assumption 2, the sub-problem defined in Section III is indexable.

While Theorem 6 proves indexability for Markovian energy arrivals, computing the Whittle Index for this case is an open problem. One way to compute the Whittle Index, also used in [14], is to show that a cost-optimal policy (Definition 2) is a stationary deterministic policy of the threshold-type. Then, using the fact that the cost is convex in this threshold, the Whittle Index is computed by solving for that value of the cost c at which idling and sending an update given the current AoI of the source is equally desirable. In this case, since there are multiple thresholds corresponding to the various battery/energy arrival states, this technique to find the Whittle Index is not effective. To address this limitation, we use a proxy for the Whittle Index [15] and use it to design a scheduling policy.

Definition 7: (Modified Whittle Index [15]) For a given tolerance parameter $\epsilon > 0$, the Modified Whittle Index is the value of c for which the approximate value functions corresponding to the two actions (idle/update) are at most ϵ apart.

Another challenge is that the number of states in the decision problem described in Section III is countably infinite. Therefore for tractability, we truncate age by a large value k to get a finite state MDP. In the truncated system, $X(t) = \min\{X(t), k\}$. Let $S^{(k)}$ denote the set of states in the truncated MDP. By [16], for finite state-action unichain MDPs, a cost-optimal policy is stationary. The relative value iteration algorithm (RVIA) [16] for the truncated MDP is described in Algorithm 2.

Algorithm 2: RVIA FOR THE k-TRUNCATED MDP

Input: $s = (x, b, \lambda) \in S^{(k)}$, action $a \in \{0, 1\}$, p, q, c, tolerance ϵ, $i_0 \in S^{(k)}$ as a reference state.

Initialize: $V_1^{(k)}(s) = 0, \forall s \in S^{(k)}$
foreach $s \in S^{(k)}$ **do**
  while *true* **do**
    $V_{n+1}^{(k)}(s) = \min_{a\in\{0,1\}} C(s, a) + \mathbb{E}[V_n^{(k)}(s')] - V_n^{(k)}(i_0)$;
    if $|V_{n+1}^{(k)}(s) - V_n^{(k)}(s)| < \epsilon$ **then**
      $V^{(k)}(s) = V_{n+1}^{(k)}(s)$.
      break;
    n=n+1;

By Theorem 8.6 in [17], we know that for a finite MDP, if all stationary policies are unichain, i.e., the Markov chain corresponding to every deterministic stationary policy consists of a single recurrent class plus a possibly empty set of transient states, then the policy iteration converges in a finite number of iterations to the average optimal stationary policy, and it satisfies the corresponding average cost optimality equation.

Next, we argue that the truncated MDP is unichain. From [17], we know that for every truncated MDP, there is only one recurrent class. Since the state $(k, 1, 1)$ is reachable from all other states, the truncated MDP in unichain for any value of k. Using Theorem 2.2 in [18], we show that the sequence of approximate MDPs indexed by k converges to the original MDP in this setting.

We now present an algorithm to compute the Modified Whittle Index for the truncated MDP. For a given $s = (x, b, \lambda) \in \mathcal{N} \times [0, 1]^2$ start at $t = 1$ with an initial cost to play c_0 and run the value iteration algorithm to compute the value functions $V^{(k)}(s, a = 1)$ and $V^{(k)}(s, a = 0)$. The cost c_0 is updated as follows: $c_{t+1} = c_t + \alpha(V^{(k)}((s, a = 1), c_t) - V^{(k)}((s, a = 0), c_t))$where α is a learning parameter. The algorithm terminates when $|V^{(k)}(s, a = 1) - V^{(k)}(s, a = 0)| < \epsilon$, where ϵ is the tolerance limit. See Algorithm 3 for a formal definition. Note that we use two timescales, one for updating the cost and the other for updating value functions. The value of the learning parameter α is chosen such that the cost c_t is updated at a slower timescale compared to the value iteration algorithm that computes $V^{(k)}((s, a = 1), c_t)$ and $V^{(k)}((s, a = 0), c_t)$. This is a two-timescale stochastic approximation and is based on similar schemes studied in [15], [19]. The convergence of two-timescale stochastic approximations is discussed in Chapter 6 of [19]. If α_t is replaced with a tiny constant value α, there is convergence with high probability as shown in [19].

Algorithm 3: MODIFIED WHITTLE INDEX COMPUTATION (k-TRUNCATED MDP)

Input: State $s = (x, b, \lambda)$, action $a \in \{0, 1\}$, initial cost c_0, step size α, tolerance ϵ.

Output: Whittle's index $I(s)$

procedure For every $s \in \mathcal{N} \times \{0, 1\}^2$;
$c_t \leftarrow c_0$
while $|V^{(k)}(s, a = 1) - V^{(k)}(s, a = 0)| > \epsilon$ **do**
  $c_{t+1} = c_t + \alpha(V^{(k)}((s, a = 1), c_t) - V^{(k)}((s, a = 0), c_t))$;
  $t = t + 1$;
  Compute $V^{(k)}((s, a = 1), c_t)$, $V^{(k)}((s, a = 0), c_t)$;
return $I(s) \leftarrow c_t$.
end procedure

Algorithm 4: MODIFIED WHITTLE-MARKOV SCHEDULING

Input: Number of sources n, (x_i, b_i, λ_i), $p_i, q_i, \forall i$.

procedure In each time-slot, compute Modified Whittle Index $I(x_i, b_i, \lambda_i), \forall i$ using Algorithm 3.
Schedule Source i^* for communication, where $i^* = \arg \max_{i\in\{1,2,\cdots,n\}} I(x_i, b_i, \lambda_i)$.
end procedure

V. PROOFS

A. i.i.d. Energy Arrivals

The proof for this setting follows the same steps as in [14]. We omit the details for this case due to space constraints.

B. Markovian Energy Arrivals

The first step towards proving Theorem 6 is to show that a cost-optimal policy (Definition 2) for the decoupled sub-problem discussed in Section III is a stationary deterministic policy.

Lemma 1: For the decoupled sub-problem discussed in Section III, under Assumption 2, there exists a stationary and deterministic policy that is cost-optimal, independent of initial state $s(0)$.

Proof: By Theorem 12 in [14], a deterministic stationary policy is cost-optimal if the following two conditions hold.

1) There exists a deterministic stationary policy of the MDP such that the associated average cost is finite.
2) There exists a non-negative L such that the relative cost function $h_\beta(s) = V_\beta(s) - V_\beta(0) \geq -L$ for all s and β, where 0 is a reference state.

We now prove the first condition. Let f be the stationary deterministic policy of choosing to update the monitoring station in every time-slot when the source battery has sufficient energy. The state $(X(t), \Lambda(t))$ under this policy f forms a discrete time Markov chain (DTMC) as shown in Figure 1. The steady state distribution $\pi = [\pi_{10}, \pi_{11}, \pi_{20}, \pi_{21}, \ldots]$ of the DTMC is given by

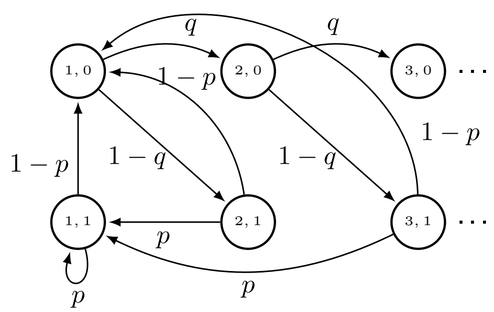

Fig. 1: The AoI $X(t)$ under the policy f forms a DTMC.

$$\pi_{10} = \frac{(1-p)(1-q)}{2-p-q}; \pi_{11} = \frac{p(1-q)}{2-p-q};$$
$$\pi_{xo} = q^{x-1}\pi_{10}, \forall x \in \{2, 3, \ldots\};$$
$$\pi_{x1} = q^{x-2}(1-q)\pi_{10}, \forall x \in \{2, 3, \ldots\}.$$

The average AoI is given by

$$\sum_{i=1}^{\infty} i(\pi_{i0} + \pi_{i1}) = \frac{(1-p)(2-q) + (1-q)^2}{(2-p-q)(1-q)}.$$

On the other hand, the average updating cost is $\frac{c(1-q)}{2-p-q}$. Hence, the average cost per time slot under this policy is

$$\frac{(1-p)(2-q) + (1-q)^2}{(2-p-q)(1-q)} + \frac{c(1-q)}{2-p-q},$$

which is finite and yields the desired result.

Next, we prove the second condition. The value function equation $V_\beta(x, b, \lambda)$ is given by

$$V_\beta(x, b, \lambda) = \min\{V_\beta\big((x, b, \lambda), a = 0\big), V_\beta\big((x, b, \lambda), a = 1\big)\},$$

We show that $V_\beta(x, b, \lambda)$ is a non-decreasing function of AoI x for a given β, b, λ. We omit the details due to space constraints. From Proposition 5 in [20], if the first condition of Theorem 1 is satisfied, then for a given state (x, b, λ) and discount factor β, the quantity $V_\beta(x, b, \lambda)$ is finite. Therefore, $k = V_\beta(1, 0, 0) - V_\beta(1, 1, 1)$ is finite. It follows that $V_\beta(x, b, \lambda) - V_\beta(1, 1, 1) \geq k; \forall (x, b, \lambda), \forall \beta$. Choose $V_\beta(1, 1, 1)$ as a reference state and choose a non-negative L appropriately based on k such that $V_\beta(x, b, \lambda) - V_\beta(1, 1, 1) \geq -L; \forall x, \forall b, \forall \lambda, \forall \beta$, thus proving the result. ∎

Next we show that a specific stationary deterministic policy, namely a threshold-type policy is cost-optimal.

Lemma 2: If $c \geq 0$, then for the decoupled sub-problem discussed in Section III, under Assumption 2, there exists a policy of the threshold type that is cost-optimal.

Proof: From the value function equations, we can see that the optimal action for state $(x, 0, 0)$ is to idle if $c \geq 0$. From [14], if the first condition in the proof of Lemma 1 holds, then for any state s the minimum expected total discounted cost $V_\beta(s)$ satisfies $V_\beta(s) = \min_{a \in \{0,1\}} C(s, a) + \beta \mathrm{E}[V_\beta(s')]$, where the expectation is taken over all possible next state s' reachable from state s.

We now prove that an β-optimal policy is the threshold type. Suppose an β-optimal action for state $(x, 1, 0)$ is to update, i.e., $V_\beta((x, 1, 0), 1) - V_\beta((x, 1, 0), 0) \leq 0$. Then, the β-optimal action for state $(x+1, 1, 0)$ is also to update since

$$\begin{aligned}
&V_\beta((x+1, 1, 0), 1) - V_\beta((x+1, 1, 0), 0) \\
=&\big(1 + c + \beta(qV_\beta(1, 0, 0) + (1-q)V_\beta(1, 1, 1))\big) \\
&- \big(x + 2 + \beta\big(qV_\beta(x+2, 1, 0) + (1-q)V_\beta(x+2, 1, 1)\big)\big) \\
\leq&\big(1 + c + \beta(qV_\beta(1, 0, 0) + (1-q)V_\beta(1, 1, 1))\big) \\
&- \big(x + 1 + \beta\big(qV_\beta(x+1, 1, 0) + (1-q)V_\beta(x+1, 1, 1)\big)\big) \\
=&V_\beta((x, 1, 0), 1) - V_\beta((x, 1, 0), 0) \leq 0,
\end{aligned}$$

where the above result is from the non-decreasing function of $V_\beta(x, b = 1, \lambda = 0)$ in x for fixed values of b, λ.

Similarly, suppose the β-optimal action for state $(x, 1, 1)$ is to update, i.e., $V_\beta((x, 1, 1), 0) - V_\beta((x, 1, 1), 0) \leq 0$. Then, it can be shown that the β-optimal action for state $(x+1, 1, 1)$ is also to update, we omit the details due to space constraints.

Hence, a β–optimal policy is a threshold type policy with two thresholds corresponds to states $(x, 1, 0)$ and $(x, 1, 1)$. Finally, we consider the limit $\beta \to 1$ and obtain the limit point of the β-optimal policies as a cost-optimal policy [20]. We thus conclude that a cost-optimal policy is also of the threshold type. ∎

Proof of Theorem 6: Let $S_1(c)$, $S_2(c)$ and $S_3(c)$ be the sets of states of the form $(x, 0, 0)$, $(x, 1, 0)$ and $(x, 1, 0)$ respectively, for which the optimal action is to idle. Let $S(c) = S_1(c) \cup S_2(c) \cup S_3(c)$. Recall that, the optimal action for state $(x, 0, 0)$ is to idle if $c \geq 0$. So corresponding set $S_1(c) = \{(x, 0, 0) : x = 1, 2, \ldots\}$. For state $(x, 1, 0)$,

$$V_\beta\big((x, 1, 0), a = 0\big) \underset{a=1}{\overset{a=0}{\lesseqgtr}} V_\beta\big((x, 1, 0), a = 1\big),$$

$$\text{i.e., } x + 1 + \beta\big(qV_\beta(x+1, 1, 0) + (1-q)V_\beta(x+1, 1, 1)\big)$$
$$\underset{a=1}{\overset{a=0}{\lesseqgtr}} 1 + c + \beta\big(qV_\beta(1, 0, 0) + (1-q)V_\beta(1, 1, 1)\big).$$

We know that $V_\beta(x, b, \lambda)$ is non-decreasing in x for a given b, λ, β. Note that $\forall c \geq 0$, $\exists x_0$ such that $x + \beta\big(qV_\beta(x + 1, 1, 0) + (1-q)V_\beta(x+1, 1, 1)\big) > c + \beta\big(qV_\beta(1, 0, 0) + (1-q)V_\beta(1, 1, 1)\big), \forall x > x_0$. Therefore, the optimal action is to idle till x_0. As c increases, x_0 increases, so the set $S_2(c) = \{(x, 1, 0) : x = 1, 2, \ldots, x_0\}$ monotonically increases. Similarly we can show that for state $(x, 1, 1), \exists x_1$ such that $S_3(c) = \{(x, 1, 1) : x = 1, 2, \ldots, x_1\}$ which increases monotonically with c. Therefore, the set $S(c)$ monotonically increases to the entire state space as $c \to \infty$ and the sub-problem is indexable by Definition 3. ∎

VI. Simulation Results

In this section, we compare the performance of Modified Whittle-Markov policy (Algorithm 4) with Whittle-iid (Algorithm 1) and the greedy policy which schedules the source with maximum weighted age in each time-slot. We refer to the greedy policy as Myopic. Since the energy arrival process is Markovian, for the Whittle-iid policy, we use the steady-state probability of an energy arrival in a time-slot as the energy arrival parameter.

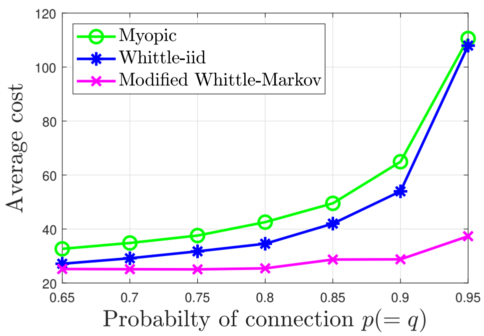

Fig. 2: Average cost as a function of connection probability for independent arrivals across sources.

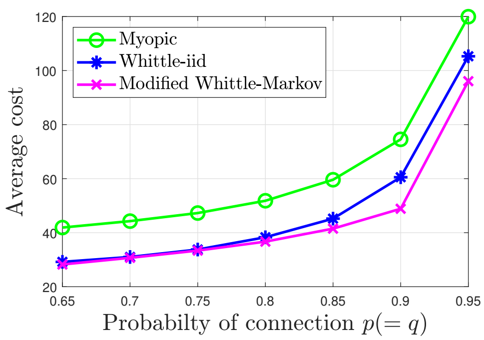

Fig. 3: Average cost as a function of connection probability for correlated arrivals.

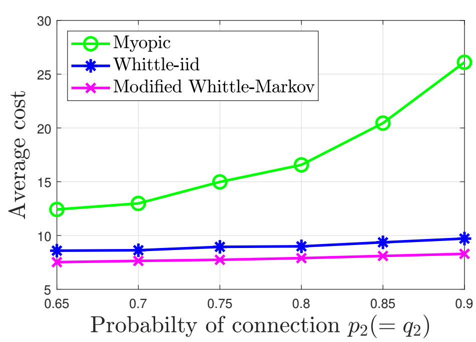

Fig. 4: Average cost as a function of connection probability.

We fix $\epsilon = 0.1$, $\alpha = 0.05$ for the Modified Whittle-Markov policy. In Figures 2 and 3, we simulate a system of four sources and vary the parameters of the energy arrival process. The weights of the sources are $[1, 2, 3, 4]$. Here $p_i = q_i = p = q$ for all i. In Figure 2, the energy arrival processes are independent across the four sources and in Figure 3 the energy arrival process is identical (sample-path wise) across sources. The Modified Whittle-Markov policy outperforms the Whittle-iid policy and Myopic policy in both cases. In Figure 4, we simulate a system of two sources. Here $p_1, q_1 = 0.7$ and we vary the value of $p_2 = q_2$. The weights of the source are $[1, 4]$. The Modified Whittle-Markov policy outperforms the Whittle-iid policy and Myopic policy.

References

[1] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *IEEE INFOCOM 2009*. IEEE, 2009, pp. 1422–1430.

[2] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "Age and value of information: Non-linear age case," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 326–330.

[3] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal online policies for energy harvesting sensors with incremental battery recharges," in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–10.

[4] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 193–204, 2017.

[5] S. Park, H. Kim, and D. Hong, "Cognitive radio networks with energy harvesting," *IEEE Transactions on Wireless communications*, vol. 12, no. 3, pp. 1386–1397, 2013.

[6] S. Leng and A. Yener, "Age of information minimization for an energy harvesting cognitive radio," *IEEE Transactions on Cognitive Communications and Networking*, 2019.

[7] S. Farazi, A. G. Klein, and D. R. Brown, "Age of information in energy harvesting status update systems: When to preempt in service?" in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2436–2440.

[8] W. Liu, X. Zhou, S. Durrani, H. Mehrpouyan, and S. D. Blostein, "Energy harvesting wireless sensor networks: Delay analysis considering energy costs of sensing and transmission," *IEEE Transactions on Wireless Communications*, vol. 15, no. 7, pp. 4635–4650, 2016.

[9] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 3008–3012.

[10] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2988–3001, 2013.

[11] V. Deulkar, J. Nair, and A. A. Kulkarni, "Sizing storage for reliable renewable integration," in *2019 IEEE Milan PowerTech*. IEEE, 2019, pp. 1–6.

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

[13] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of applied probability*, vol. 25, no. A, pp. 287–298, 1988.

[14] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Transactions on Mobile Computing*, 2019.

[15] K. Kaza, R. Meshram, V. Mehta, and S. N. Merchant, "Sequential decision making with limited observation capability: Application to wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, 2019.

[16] J. Abounadi, D. Bertsekas, and V. S. Borkar, "Learning algorithms for markov decision processes with average cost," *SIAM Journal on Control and Optimization*, vol. 40, no. 3, pp. 681–698, 2001.

[17] M. L. Puterman, *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[18] L. I. Sennott, "On computing average cost optimal policies with application to routing to parallel queues," *Mathematical methods of operations research*, vol. 45, no. 1, pp. 45–62, 1997.

[19] V. Borkar, "Stochastic approximation: a dynamical systems view," *Hindustan Publ. Co., New Delhi, India and Cambridge Uni. Press, Cambridge, UK*, 2008.

[20] L. I. Sennott, "Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs," *Operations Research*, vol. 37, no. 4, pp. 626–633, 1989.