Computer Networks and Communications
https://ojs.wiserpub.com/index.php/CNC

Article

Whittle Index Based Age-of-Information Aware Scheduling for Markovian Channels

Bejjipuram Sombabu[1], Bhishma Dedhia[2], and Sharayu Moharir[3,*]

[1] R&D team, Mavenir Systems Pvt Ltd, Bengaluru, India
[2] Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ, USA
[3] Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
E-mail: sharayum@ee.iitb.ac.in

This paper is an extended version of our paper published in Sombabu, B.; Mate, A.; Manjunath, D.; Moharir, S. Whittle Index for AoI-Aware Scheduling. In Proceedings of 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengaluru, India, 7–11 January 2020.

Received: 16 July 2022; **Revised:** 12 December 2022; **Accepted:** 12 December 2022

Abstract: The focus of this work is on minimizing the time average of the weighted sum of the Age-of-Information in a multi-sensor system in the setting where all sensors report their measurements to a central monitoring station using a shared communication channel. We consider multiple information settings including complete CSI/delayed CSI/no CSI and two stochastic channel evolution models, i.e., i.i.d. channels and Markovian channels. In all settings considered, we prove the indexability of the scheduling problem. In addition, we compute the Whittle Index in closed form for some of the settings. Indexability for Markovian channels for the objective of minimizing AoI is a key contribution of this work. Further, under i.i.d. channels, we propose a novel efficient implementation for the Whittle Index-based policy. We use simulations to show that Whittle Index-based scheduling policies either outperform or match the performance of the state-of-the-art policies for all the settings considered.

***Keywords*:** age-of-information, scheduling, communication channel

1. Introduction

The age-of-information (AoI) metric [1] is a performance metric introduced about a decade ago and is a measure of the freshness of information available at the intended destination of the information. Formally, the age-of-information is defined as the time elapsed since the most recent update packet at the receiver was generated at the source. This metric has spurred a significant amount of research activity in areas like scheduling [2–9], caching [10–12], energy harvesting [13–15], and channel state information estimation [16–18].

This work is motivated by the Internet of Things (IoT) which has applications across domains like infrastructure, security, and medicine. A typical IoT network has a large number of sensors tracking/measuring physical phenomena and reporting their findings to a central monitoring station. The monitoring station typically uses the status updates received from the sensors to make control decisions. The efficacy of these control decisions depends heavily on the freshness of the information used to make these decisions. This makes the AoI metric suitable to evaluate the performance of scheduling policies in IoI-type networks.

We consider a network consisting of n sensors sending updates to a monitoring station using a shared communication channel. We consider slotted time and at most one sensor can use the shared communication channel in each slot. The algorithmic challenge is to determine which sensor will communicate with the

DOI: https://doi.org/10.37256/cnc.1120231701

monitoring station in each time-slot. The performance metric is the weighted sum of the AoIs of the sensors where the weights are given to us as inputs. This problem has also been studied in [6–8] and we explore the use of Whittle Index-based policies for this setting. This is an extended version of [19].

This scheduling problem is a Restless Multi-arm Bandit problem (RMAB) by since we have sensors that can be viewed as arms with their AoI evolving in each time-slot, even when they are not scheduled for transmission. The RMAB problem is PSPACE-hard [20]. In [21], Whittle focused on a relaxed version of the RMAB problem by replacing the hard constraint on the number of arms that can be pulled in a time-slot with a constraint on its expected value. The relaxed problem is analytically tractable since it decouples across arms. The solution to the relaxed problem is known as the Whittle Index policy. The Whittle Index policy is known to perform well for some AoI-based scheduling problems [5,8,9]. A key point to note is that not all RMABs have a Whittle Index policy and those that do are called *indexable*. The contributions of this work can be summarized as follows.

1.1 *Our Contributions*

We consider ON-OFF channels. We focus on two specific channel models: (i) independent channel realizations across time-slots and (ii) Markovian channel realizations across time-slots. Channel evolution is independent across users in all cases considered. We consider multiple channel state information (CSI) settings. Specifically, we show that four instances of the problem (independent channels without CSI, independent channels with CSI, Markovian channels with CSI and Markovian channels with delayed CSI) are indexable. Further, we characterize the Whittle Index on closed-form for the first three cases. Proving indexability is often challenging and while some settings with i.i.d. channels for AoI aware scheduling are known to be indexable [5,8,9]. For the setting where channel realizations are i.i.d. across time, the Markov chain modeling the evolution of the AoI of each sensor is one dimensional and computing its stationary distribution is straightforward. For Markovian channels, the current channel state also needs to be captured in the state of the Markov chain modeling the evolution of the AoI of each sensor. This increases the state-space of the Markov chain and its structure makes computing the stationary distribution complicated. We leverage the fact that the state transition matrix is diagonalizable to compute the stationary distribution, proving indexability for Markovian channels for AoI minimization is non-trivial and is a key contribution of this work.

We compare the performance of the Whittle Index-based scheduling policies with other state-of-the-art policies through simulations. We show that our Whittle Index-based policies either outperform or match the performance of the best-known policy for all scenarios considered. In addition to this, for the setting with independent channels without CSI we propose efficient implementations of the Whittle Index type policies.

1.2 *Related Work*

As discussed above, there is a rich body of work that uses the AoI metric for various applications [2–8], [10–17], [22–29]. Refer to [30] for a comprehensive survey of results in this area.

In [8], the authors propose an index-based resource allocation policy for stochastic arrivals in the case when the realization of this stochastic process for a given time-slot is known a priori. This maps to our setting of i.i.d. channels with known CSI. In [9], the authors propose a Whittle Index-based policy for AoI minimization for i.i.d. channels in the setting where there are throughput constraints. Unlike [9], there are no throughput constraints in this work. In addition, compared to [9], we show that Whittle Index-based policies are effective for more general channel models. In [5], the authors provide analytical performance guarantees for Whittle Index-based scheduling policies for AoI minimization for i.i.d. channel realizations in a system with two sources. In [31], the focus is on AoI-aware scheduling for specific arrival patterns like periodic arrivals and i.i.d. stochastic arrivals. This work uses two channel models, the first where the channels are reliable and the second where the channels are unreliable with i.i.d. realizations across time. In [32], the focus is on AoI-aware scheduling for broadcast communications with channels are unreliable with i.i.d. realizations across time. Further, [33] considers the scheduling problem for a system with multiple users to minimize the average of a variant of the AoI metric and establishes asymptotic optimality of the Whittle Index-based policy for i.i.d. channels. The setting considered in [34] also looks at AoI minimization for i.i.d. channels and is considerably different to the setting in our paper. In [34] time is divided into frames and each frame consists of multiple time-slots. Each source generates one packet in a frame. Old packets are discarded once a new packet is generated by a source. The authors propose three low-complexity transmission scheduling policies and evaluate their performance against the optimal policy. These policies are a randomized policy, a MaxWeight policy and a Whittle-Index based policy. The key conclusions in this paper are that in contrast to their theoretical results, simulation results show that Max-Weight and Whittle-Index based policies outperform the other policies. In [22], the authors consider a multi-source system where each source has its own intended destination. Packets/updates are generated at the source according to a stochastic process and are forwarded to a base-station that sends them to the intended destinations. Updates vary in size and their size is determined by a stochastic process. The goal is to minimize the AoI. Channel realizations are assumed to be i.i.d. across time. Compared to these recent works, our work considers the more general setting of Markovian channel evolution. In [28], the system consists of energy harvesting sources and energy arrivals to these sources as modeled as a Markov process. When a new quantum of energy arrives, it can be stored in a battery and used in subsequent time-slots. As a result, the structure of the MDP and the technical challenges in using the Whittle Index are different from that in the case of Markovian channels.

AoI-aware schedule has also been modeled as an MDP/optimization problem without using the Whittle Index to solve it. For instance, in [23], a multi-source setting with i.i.d. channels is considered. The goal is again to minimize the AoI at the intended destination when packet/update generation at the source is a stochastic process. Channels are assumed to be i.i.d. in this work. The scheduling problem is modeled as an MDP. In [24], the authors consider the setting where a base-station receives time-sensitive data from multiple sensors via error-prone Markovian channels. The data freshness is modeled as a monotonically increasing function of AoI. The goal is to design a scheduling policy to maximize data freshness in the presence of bandwidth and power constraints. Unlike [24], we consider AoI as the metric of interest and do not have any power constraints. In [25], the authors study a multi-source multi-destination system where requests for the data generated by the sources comes from the destination. In this work, sources/sensors are energy constrained and the metric of interest is a variant of AoI. The network has an edge server equipped with a cache that can store information and forward it to the destination instead of fetching fresh updates from the sources. In [26], the authors focus on a multi-source system. Here, power control is used to ensure that the capacity of the channel is large enough to ensure that the packet is received without error for all channel states. Further, in [26], the sensors have an average power consumption constraint while our work has no such constraint and instead focuses on ON-OFF channels without any power control. In [27], the goal is to minimize AoI in a system consisting of a single source that sends updates to the intended destination via a server node. The service process at the server node is modeled as a Gilbert-Elliot process with a good and a bad state. The key difference between our work and [27] is that unlike [27] which has a single source, we have multiple sources that share one communication channel to send updates to the destination.

2. Setting

We study a system of n sensors using one shared communication channel to report their measurements to a central monitoring station. We consider the case where each sensor measures an independent physical quantity.

2.1 *Network Model*

We consider slotted time. At most one sensor can communicate with the monitoring station in each time-slot. The sensors are modeled as active sources that can generate fresh updates when needed. Since the goal is to minimize the AoI at the monitoring station, when a sensor is scheduled for transmission, the sensor generates a fresh update and attempts to send it to the monitoring station.

2.2 *Channel Models*

We focus on ON-OFF channels where the communication between a sensor and the monitoring station is successful if the sensor sends an update to the monitoring station when the channel between the sensor and the monitoring station is ON and is unsuccessful otherwise.

Assumption 1: (i.i.d. ON-OFF Channels) Channel realizations are independent across sensors, and independent and identically distributed (i.i.d.) across time-slots. Let p_i denote the probability that the channel between Sensor i and the monitoring station is ON in a time-slot and $\Lambda_i(t)$ capture the state of the channel between Sensor i and the monitoring station in time-slot t, where $\Lambda_i(t) = 1$ implies that the channel is ON and $\Lambda_i(t) = 0$ denotes that the channel is OFF. It follows that

$$\mathrm{P}(\Lambda_i(t) = 1) = p_i$$

$$\mathrm{P}(\Lambda_i(t) = 0) = 1 - p_i$$

Assumption 2: (Markovian ON-OFF Channels) In this model, channel realizations are correlated across time, but independent across sensors. For each sensor, the channel state evolves as a two state (DTMC) (Figure 1). Let $\Lambda_i(t)$ denote the state of the channel between Sensor i and the monitoring station in time-slot t. Then, for Markovian channels, we have that,

$$\mathrm{P}(\Lambda_i(t+1)=1 \mid \Lambda_i(t)=1)=p_i$$

$$\mathrm{P}(\Lambda_i(t+1)=0 \mid \Lambda_i(t)=0)=q_i$$

$$\mathrm{P}(\Lambda_i(t+1)=0 \mid \Lambda_i(t)=1)=1-p_i$$

$$\mathrm{P}(\Lambda_i(t+1)=1 \mid \Lambda_i(t)=0)=1-q_i$$

Note that this channel model is known as the Gilbert-Elliott channel.

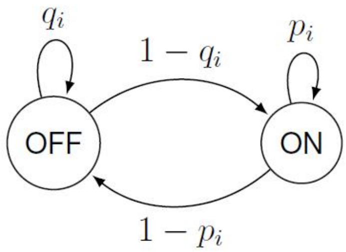

Figure 1. The Gilbert-Elliot channel model for Sensor i.

2.3 *Age of Information Model*

Let $X_i(t)$ denote the AoI of Sensor i at the monitoring station at the beginning of time-slot t. Since we assume that all sensors are active sources that can generate updates at will, and generate a new update to send to the monitoring station when scheduled for transmission, it follows that $X_i(t)$ is equal to the number of time-slots elapsed since the monitoring station received the recent most update from Sensor i. Let $D(t)$ denote the index of the sensor scheduled for communication with the monitoring station in time-slot t. For both i.i.d. and Markovian channels, the evolution of the AoI for Sensor i is as follows:

$$X_i(t+1)=\begin{cases}1 & \text{if } D(t)=i \text{ and } \Lambda_i(t)=1; \\ X_i(t)+1 & \text{otherwise}.\end{cases}$$

Remark: 1: Note that only a fraction of all attempted communications by each sensor are successfully received by the monitoring station since channels can be OFF leading to a communication failure.

2.4 *Problem Formulation*

Our goal is to minimize the time average of the weighted sum of AoIs, i.e.,

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

where, ω_i denotes the weight of Sensor i and is given to us. Here, E_θ denotes the expected value of the weighted sum of AoIs under the scheduling algorithm θ. The algorithmic challenge, therefore, is to find the θ that minimizes the time average of the weighted sum of AoIs.

3. Prelimiaries: Whittle Index

Recall that the relaxed version of the RMAB problem decouples into n sub-problems, one for each sensor [21] where each sub-problem consists of a single sensor, one channel, and the monitoring station. In addition to these, each decoupled sub-problem has an additional cost c (Lagrange multiplier) incurred for each scheduled transmission from the sensor to the monitoring station. Solving each sub-problem is equivalent to deciding whether a sensor is to be scheduled for transmission in each time-slot. In the rest of this section, we omit the sensor index i as each sub-problem deals with exactly one sensor.

Each sub-problem is a Markov decision process (MDP) [35], with the following components.

1. States: The state of the sensor is at time t is denoted by $s(t) \in \mathrm{Z}^{+}$.
2. Actions: We use $a(t) \in \{0,1\}$ to denote the action taken in time-slot t. Here, $a(t)=1$ denotes that the central scheduler attempts to send an update to the monitoring station from the sensor in time-slot t and $a(t)=0$ if the sensor remains idle in time-slot t.
3. Transition probabilities: Let $\mathrm{P}[s' | s, a]$ denote the transition probability from state s to state s' under action $a(t)=a$.
4. Cost: Let the state of the sensor be $s(t)$, $a(t)$ denote the action taken in time-slot t, the AoI of the sensor in time-slot t be denoted by $X(t)$, and the channel between the sensor and the monitoring station be denoted by $\Lambda(t)$. The cost incurred in time-slot t is denoted by $C(s(t), a(t))$, where

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

Here, the first term denotes the weighted AoI in the next time-slot and the second term is the cost incurred to send an update from the sensor.

A policy $\mu = \{a(1), a(2), \ldots\}$ determines the action taken over time. We say that a policy is stationary if $a(t) = a(t')$ for $s(t) = s(t')$ for all t, t'. We say that a policy is deterministic if chooses an action with certainty, i.e., with probability 1. A policy μ is said to be cost-optimal if it minimizes the average cost defined as follows

$$\lim_{T\to\infty} \sup \frac{1}{T} \mathrm{E}_{\mu}\left[\sum_{t=1}^{T} C(s(t), a(t))\right]. \tag{2}$$

Definition 2: (Indexability [36]) Given cost c, let $S(c)$ be the set of states for which the optimal action is to idle. The sub-problem is said to be indexable if the set $S(c)$ is monotonically increasing, specifically, the set starts out being empty for $c = -\infty$ and covers the entire state space as c increases to ∞.

Definition 3: (Whittle Index [36]) The Whittle Index is defined as the cost c that makes both actions for state s, i.e., to attempt transmission or to idle, equally desirable.

4. Main Results and Discussion

In this section we state and discuss our key results. The proofs of these results are discussed in Section 5.

4.1 *i.i.d. Channels without CSI*

We first present our analytical results for i.i.d. channels (Assumption 1) in the setting without CSI, i.e., the case where the state of the channel is unknown to the scheduler before it makes its scheduling decision for that time-slot. We assume that the state of the channel becomes known to the schedule by the end of the time-slot.

As discussed in Section 3, we consider the decoupled sub-problem with single sensor.

We formulate the sub-problem as a Markov decision process (MDP) [35], with the components as follows:

1. State: The state of a user at time t is equal to its AoI, $X(t)$. Thus, $s(t) = X(t)$.
2. Transition probabilities: The transition probability from state $s = x$ to state s' when action $a(t) = a$ is taken is given by

$$\mathrm{P}[s' = x+1 | s = x, a = 0] = 1;$$

$$\mathrm{P}[s' = x+1 | s = x, a = 1] = 1 - p;$$

$$\mathrm{P}[s' = 1 | s = x, a = 1] = p.$$

Theorem 4: For i.i.d. channels (Assumption 1), if scheduling decisions are made without CSI, the sub-problem is indexable and the Whittle Index for state x is given by

$$I(x) = \omega\left(\frac{px^2}{2} - \frac{px}{2} + x\right).$$

Using Theorem 4 we design a Whittle Index-based scheduling policy as follows. In each time-slot, the sensor with the highest Whittle Index is chosen to send an update to the monitoring station. We call this policy the *Whittle-iid w/oCSI* policy. We also simulate two existing policies. Under the first policy called the *Myopic w/oCSI* policy schedules the sensor with the maximum product of the weighted AoI and the probability of the channel being ON($\arg\max_i p_i \omega_i X_i(t)$) in each time-slot. The other policy called the *Myopic-modified w/oCSI* policy schedules the sensor with the maximum product of the weight and the square of AoI and the probability of the channel being ON($\arg\max_i p_i \omega_i X_i^2(t)$) in each time-slot. This policy was first proposed in [3] and was shown to perform well. In Figure 2, we vary the number of sensors. The weights of the sensors are chosen uniformly at random in [1, 20]. The $p_i s$ are chosen uniformly at random in [0.2, 1]. We observe that the *Whittle-iid w/oCSI* policy performs better than the other two policies. The average cost increases with the number of sensors as the fraction of sensors that can be served in a time-slot decreases, thus leading to higher AoI values.

In Figure 3, we simulate a system of five sensors and vary the channel statistics for one of the five sensors. Here we use the following parameter values: $p_1 = 0.4$, $p_2 = 0.5$, $p_3 = 0.6$, $p_4 = 0.3$, and p_5 is varied from 0.4 to 0.9. The weights corresponding to the five sensors are chosen uniformly at random in [1, 20]. The *Whittle-iid w/oCSI* policy outperforms the other two policies. As p_5 increases, the probability of successful transmissions increases, and therefore the average cost decreases across policies.

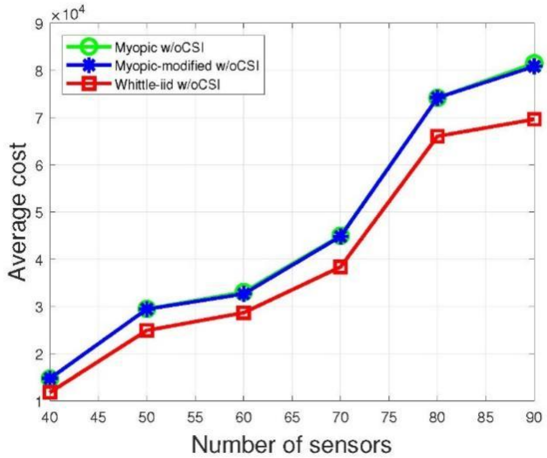

Figure 2. Average cost as a function of system size for the i.i.d. channels without CSI setting.

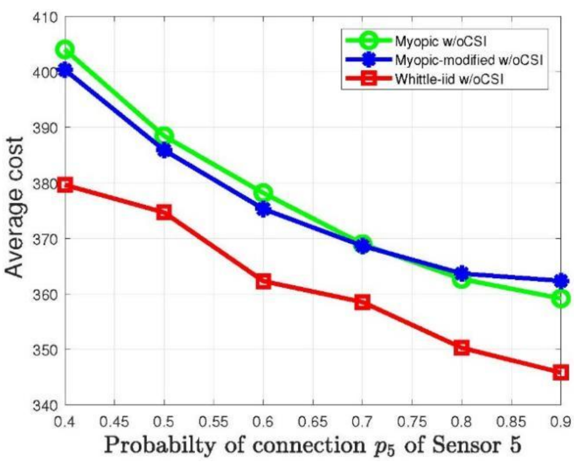

Figure 3. Average cost for a system with five sensors by varying connection probability of Sensor 5 for the i.i.d. channels without CSI setting.

In Appendix A, we describe an efficient implementation of the Whittle Index scheduling policy for large-scale systems with i.i.d. channels without CSI.

4.2 *Markovian Channels with CSI*

We now focus on the setting where the channel evolution is Markovian (Assumption 2).
We define each sub-problem as follows.

a). State: The state of a sensor at time t is defined by $s(t) = (X(t), \Lambda(t))$.

b). Transition probabilities: The transition probability from state $s = (x, \lambda)$ to state s' under action $a(t) = a$ is:

$$\mathrm{P}[s' = (x+1, 0) \,|\, s = (x, 0), a] = q;$$

$$\mathrm{P}[s' = (x+1, 1) \,|\, s = (x, 0), a] = 1 - q;$$

$$\mathrm{P}[s' = (x+1, 1) \,|\, s = (x, 1), a = 0] = p;$$

$$\mathrm{P}[s' = (x+1, 0) \,|\, s = (x, 1), a = 0] = 1 - p;$$

$$\mathrm{P}[s' = (1, 0) \,|\, s = (x, 1), a = 1] = 1 - p;$$

$$\mathrm{P}[s' = (1, 1) \,|\, s = (x, 1), a = 1] = p.$$

Theorem 5: For Markovian channels (Assumption 2), if scheduling decisions are made with CSI, the sub-problem is indexable and the Whittle Index for state (x, λ) is given by

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

where,

$$\begin{aligned} A = \omega((q^3 + (2p-5)q^2 + (p^2 - 6p + 8)q - p^2 + 4p - 4)x^2 + (q^3 + (2p-5)q^2 + (p^2 - 8p + 10)q - 3p^2 + 10p - 8)x \\ + (q + p - 1)^x((2p-2)q + 2p^2 - 4p + 2) + (2 - 2p)q - 2p^2 + 4p - 2) \end{aligned}$$

$$B = (2q^3 + (4p - 10)q^2 + (2p^2 - 12p + 16)q - 2p^2 + 8p - 8).$$

Using Theorem 5, we use the Whittle Index to design the following scheduling policy. In each time-slot, the scheduler selects the sensor with the highest Whittle Index to send an update to the monitoring station. We refer to this policy as the *Whittle-Markov w/CSI* policy. We also simulate two other policies. The *Myopic w/CSI* policy schedules the sensor with maximum weighted age ($\arg\max_i \omega_i X_i(t)$) in each time-slot. The *Myopic-modified w/CSI* policy schedules the sensor with the maximum product of weight and the square of the AoI ($\arg\max_i \omega_i X_i^2(t)$) [3].

In Figure 4, we vary the number of sensors. The weights of the sensors are chosen uniformly at random in [1, 20]. The transition probabilities p_i, q_i are also chosen uniformly at random in [0.5, 1]. We note that our Whittle-Markov w/CSI policy outperforms other policies.

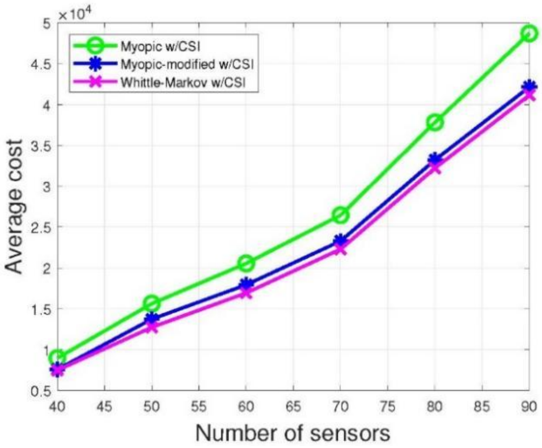

Figure 4. Average cost as a function of system size for the Markovian channels with CSI setting.

In Figure 5, we simulate a system of five sensors and vary the channel statistics for one of the sensors. For this set of simulations, we set the parameter values such that $p_i < 1 - q_i$, for all i. The parameter values are $p_1 = 0.4$, $q_1 = 0.5$, $p_2 = 0.5$, $q_2 = 0.4$, $p_3 = 0.6$, $q_3 = 0.3$, $p_4 = 0.7$, $q_4 = 0.2$, and p_5, q_5, are varied from 0.4 to 0.9 and 0.55 to 0.05 respectively. The weights corresponding to the five sensors are chosen uniformly at random in [1, 20]. Our Whittle-Markov w/CSI policy outperforms other policies.

Remark: 6: Note that the setting with i.i.d. channels with CSI is a special case of Markovian channels with CSI.

Corollary 7: For Markovian channels (Assumption 1), if scheduling decisions are made with CSI, the sub-problem is indexable and the Whittle Index for state (x, λ) is given by

$$I(x, \lambda) = \begin{cases} 0 & \text{if } \lambda = 0; \\ \omega\left(\dfrac{x^2}{2} - \dfrac{x}{2} + \dfrac{x}{p}\right) & \text{if } \lambda = 1. \end{cases}$$

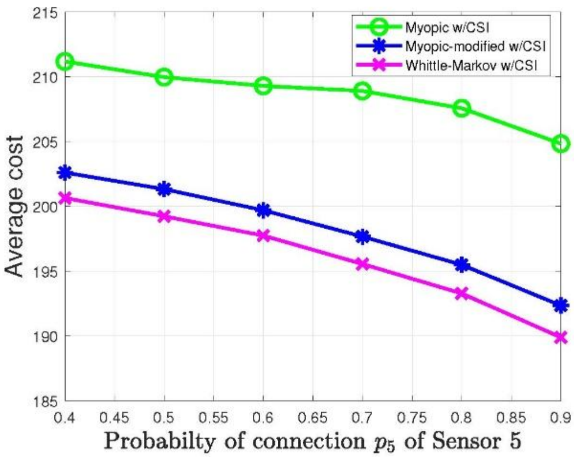

Figure 5. Average cost for a system with five sensors by varying connection probability of Sensor 5 for the Markovian channels with CSI setting.

Proof: Since the i.i.d channels are a degenerate case Markovian channels with transition probability to next state being independent of the current state. So substitute $q = 1 - p$ in Whittle Index expression in Theorem 5 to get the result.

We use Corollary 7 to use the Whittle Index to design the following scheduling policy. In each time-slot, the scheduler selects the sensor with the highest Whittle Index to send an update to the monitoring station. We refer to this policy as the *Whittle-iid w/CSI* policy. We also simulate three other policies. The first two are the Myopic w/CSI policy and the Myopic-modified w/CSI [3] defined in the previous sub-section. Further, for the setting with reliable channels, i.e., if all channels are always ON, we also simulate the *SQRT-Weight* policy proposed in [6]. This policy is known to be 8-optimal for the problem defined in Section 2. We note that there is a known lower bound (derived in [6]) on the performance of any scheduling policy when all channels are always in the ON state, i.e., $p_i = 1$ for all i. We also plot this lower bound in all applicable situations. As the number of sensors increases, the average cost increases since there is a constraint on the number of sensors scheduled per time-slot.

In Figure 6, we vary the number of sensors. Here, $p_i = 1, \forall i.$ Weights of the sensors are chosen at uniformly at random between 1 and 20. The Whittle Index based policy matches the performance of the Myopic-modified w/CSI policy and outperforms the other two policies.

In Figure 7, we simulate a system of five sensors by varying the connection probability of one of the sensors. For this set of simulations, $p_1 = 0.3, p_2 = 0.6, p_3 = 0.5, p_4 = 0.7$ and p_5 is varied from 0.4 to 0.9. The weights corresponding to the five sensors are chosen uniformly at random between 1 and 20. The Whittle Index based policy outperforms the other two policies. As p_5 increases the chances of successful transmissions increases and therefore the average cost decreases for all policies.

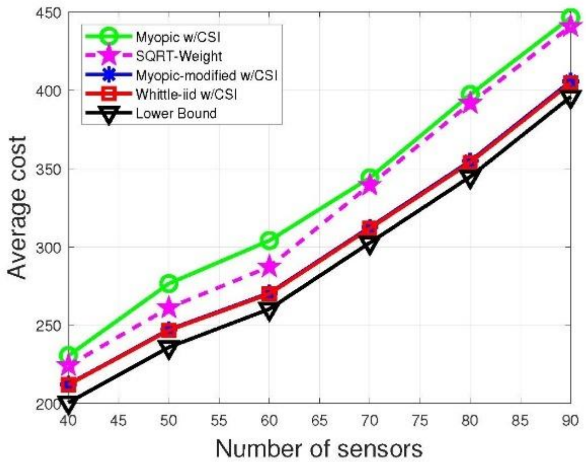

Figure 6. Average cost as a function of system size for the i.i.d. channels with CSI setting.

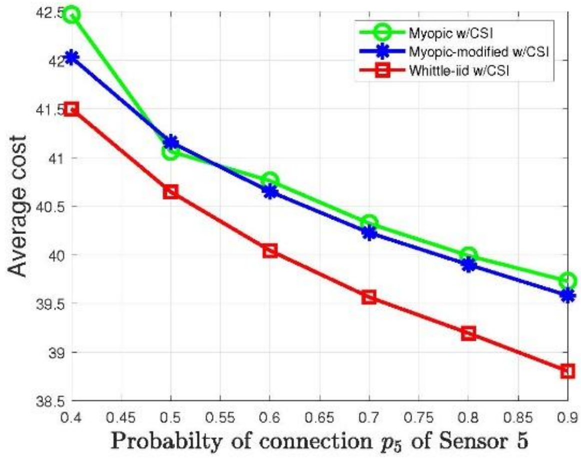

Figure 7. Average cost for a system with five sensors by varying connection probability of Sensor 5 for the i.i.d. channels with CSI setting.

4.3 *Markovian Channels with Delayed-CSI*

In this subsection, we consider the setting where the current channel state is unknown to the scheduler, however, it knows the channel state D time-slots old from the current time-slot. $\Lambda(t-D)$ represents the channel state which D time-slots old. The delay D can be any finite integer value.

We again consider the sub-problem in which we consider a single node.

We define the structure of each individual sub-problem in the following fashion.

1. *State*: We define the state of a user at time t by $s(t) = (X(t), \Lambda(t-D))$, where $\Lambda(t-D)$ is D time-slots past channel state.

2. *Transition probabilities*: The transition probability from state $s=(x,\lambda)$ to state s' under action $a(t)=a$ is:

$P[s'=(x+1,0)\,|s=(x,0),a=0]=q$;

$P[s'=(x+1,1)\,|s=(x,0),a=0]=1-q$;

$$P[s'=(x+1,0)\,|s=(x,0),a=1]=q\left[\frac{(1-p)+(1-q)(p+q-1)^{D-1}}{2-p-q}\right]:=qP_{00}(D-1);$$

$$P[s'=(x+1,1)\,|s=(x,0),a=1]=(1-q)\left[\frac{(1-p)-(1-p)(p+q-1)^{D-1}}{2-p-q}\right]:=(1-q)P_{10}(D-1);$$

$$P[s'=(1,0)\,|s=(x,0),a=1]=q\left[\frac{(1-q)-(1-q)(p+q-1)^{D-1}}{2-p-q}\right]:=qP_{01}(D-1);$$

$$P[s'=(1,1)\,|s=(x,0),a=1]=(1-q)\left[\frac{(1-q)+(1-p)(p+q-1)^{D-1}}{2-p-q}\right]:=(1-q)P_{11}(D-1);$$

$P[s'=(x+1,0)\,|s=(x,1),a=0]=1-p$;

$P[s'=(x+1,1)\,|s=(x,1),a=0]=p$;

$$P[s'=(x+1,0)\,|s=(x,1),a=1]=(1-p)\left[\frac{(1-p)+(1-q)(p+q-1)^{D-1}}{2-p-q}\right]:=(1-p)P_{00}(D-1);$$

$$P[s'=(x+1,1)\,|s=(x,1),a=1]=p\left[\frac{(1-p)-(1-p)(p+q-1)^{D-1}}{2-p-q}\right]:=pP_{10}(D-1);$$

$$P[s'=(1,0)\,|s=(x,1),a=1]=(1-p)\left[\frac{(1-q)-(1-q)(p+q-1)^{D-1}}{2-p-q}\right]:=(1-p)P_{01}(D-1);$$

$$P[s'=(1,1)\,|s=(x,1),a=1]=p\left[\frac{(1-q)+(1-p)(p+q-1)^{D-1}}{2-p-q}\right]:=pP_{11}(D-1).$$

Theorem 8: For channels satisfying Assumption 2, if scheduling decisions are made with delayed CSI, the sub-problem is indexable.

Computing the Whittle Index for this setting remains an open problem.

Remark: 9: Please refer to Appendix B for the proofs and the Indexability result for the Markovian channels with one Delayed-CSI which is a special case with $D=1$.

5. Proofs

We first provide an outline of the proof for the settings discussed in Section 4. Our proofs use techniques developed in [8] for i.i.d. channels to prove indexability and compute the Whittle Index. The proofs for the i.i.d. channel case, both with and without CSI are analytically similar to those in [8]. Our main analytical contribution is generalizing the techniques used in [8] for Markovian channels.

One of the key steps in proving indexability and computing the Whittle Index is to model the evolution of the AoI of each sensor under a given scheduling policy as a Markov chain and computing its stationary distribution. For the setting where channel realizations are i.i.d. across time, the Markov chain modeling the evolution of the AoI of each sensor is one dimensional and computing its stationary distribution is straightforward. For Markovian channels, the current channel state also needs to be captured in the state of the Markov chain modeling the evolution of the AoI of each sensor. This increases the state-space of the Markov chain and its structure makes computing the stationary distribution complicated. We leverage the fact that the state transition matrix is diagonalizable to compute the stationary distribution.

5.1 *Proof Outline*

We use the following steps to prove our results.

To characterize the cost optimal policy, we use the concept of β -discounted cost [35], where $0<\beta<1$ is the discount factor. Given the initial state $s(1)=s$, we define the expected total β -discounted cost under policy μ as

$$V_\beta(s;\mu)=\limsup_{T\to\infty} \mathrm{E}_\mu\left[\sum_{t=1}^{T}\beta^t C(s(t),a(t)\,|s(1)=s)\right].$$

For any state s, the minimum expected total discounted cost $V_\beta(s)$ satisfies

$$V_\beta(s)=\min_{a\in\{0,1\}} C(s,a)+\beta\,\mathrm{E}\left[V_\beta(s')\right],$$

where the expectation is taken over all possible next state s' reachable from state s. We begin by showing that a deterministic stationary policy is cost-optimal. To prove this, we verify if the two conditions in the following lemma hold (Theorem 12 in [8]).

Lemma 1: 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 average cost of the resulting Markov chain 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 denotes the reference state.

Next we show that a cost-optimal policy is a special type of deterministic stationary policy, which is a threshold type policy. To prove this, we first prove that an optimal policy for the β -discounted cost problem is of the threshold type. Note that a policy μ is optimal for the β -discounted cost problem if $V_\beta(s)=\min_\mu V_\beta(s;\mu)$.

The MDP defined in Section 3 is unichain. Given this, the cost optimal policy is the limit point of the β -optimal policy [35]. We use this to conclude that a cost-optimal policy is of the threshold type. Formally, we prove the following result.

Lemma 2: If $c\geq 0$, then there exists a policy of the threshold type that is cost-optimal.

We show that the problem is indexable by proving that the set of states for which the optimal action is to idle monotonically increases to the entire state space as the cost $c\to\infty$ (Definition 2).

Next, for each of the first three settings discussed in [4], we characterize the average cost for a threshold type policy as a function of the threshold $\bar{X}$ and the cost c for playing an arm. We denote this cost by $C_{avg}(\bar{X},c)$. We show that $C_{avg}(\bar{X},c)$ is a convex function of the threshold $\bar{X}$. Let $x^*(c)$ (not necessarily an integer) be the minimizer of $C_{avg}(x,c)$ for a given c. Using the convexity of the cost function, we have that, the optimal threshold $\bar{X}$ is either $\lfloor x^*(c)\rfloor$ or $\lceil x^*(c)\rceil$. As mentioned in Definition 3, Whittle Index at a given state is defined as the value of cost c at which both actions (update/not update) are equally optimal. Both actions will be equally optimal at $\lfloor x^*(c)\rfloor$ if and only if the average cost with $\lfloor x^*(c)\rfloor$ as the threshold is equal to the average cost with $\lceil x^*(c)\rceil$ as the threshold. Given this, for a given age-of-information x, the Whittle Index is equal to the value of c which satisfies $C_{avg}(x,c)=C_{avg}(x+1,c)$.

5.2 *i.i.d. Channels without CSI*

Proof of Lemma 1:

1. We let f be the stationary deterministic policy of choosing to update the user in every time-slot. Under this policy, for a node with an AoI of $X(t)=x$, the AoI becomes 1 in the next slot with probability p and increases to $x+1$ with probability $(1-p)$. The AoI $X(t)$ under policy f is a discrete-time Markov chain (DTMC) (Figure 8). The steady state distribution $\pi=[\pi_1,\pi_2,...]$ of the DTMC is

$$\pi_i = p(1-p)^{i-1},$$

for all $i=1,2,....$ Hence, the average age is

$$\sum_{i=1}^{\infty} i\pi_i = \sum_{i=1}^{\infty} ip(1-p)^{i-1} = \frac{1}{p}.$$

Recall that the updating cost is c. Hence, the average cost per time slot under this policy is $(1/p)+c$ which is finite, thus yielding the result.

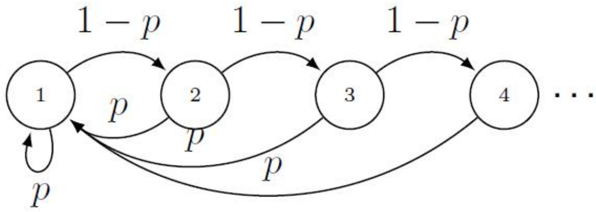

Figure 8. The age $X(t)$ under the policy f forms a DTMC.

2. For our problem $V_\beta(x)$ is given by

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

where, $V_\beta(x,a=0)=x+1+\beta V_\beta(x+1)$,

and $V_\beta(x,a=1)=p(1+c)+(1-p)(x+2+c)+\beta\left(pV_\beta(1)+(1-p)V_\beta(x+2+c)\right)$.

From above, $V_\beta(x)$ is a non-decreasing function of AoI x for a given β. From Proposition 5 of [37] if the first condition of Lemma 1 satisfies then $V_\beta(x)$ is finite for any given x, β. Consider state 1 as reference state and choose $L=0$, then $V_\beta(x)-V_\beta(1)\geq 0, \forall x.$ By verifying the two conditions, the lemma immediately follows.

Proof of Lemma 2: Suppose a β-optimal action for state x is to update, i.e.,

$$V_\beta(x,a=1)-V_\beta(x,a=0)<0.$$

Then, a β-optimal action for state $x+1$ is also to update since:

$$\begin{aligned}
&V_\beta(x+1,a=1)-V_\beta(x+1,a=0)\\
&=p(1+c)+(1-p)(x+2+c)+\beta\left(pV_\beta(1)+(1-p)V_\beta(x+2)\right)-\left(x+2+\beta V_\beta(x+2)\right)\\
&=\left(p(1+c)+\beta pV_\beta(1)+c(1-p)\right)-\left(p(x+2+\beta V_\beta(x+2))\right)\\
&\leq\left(p(1+c)+\beta pV_\beta(1)+c(1-p)\right)-\left(p(x+1+\beta V_\beta(x+1))\right)\\
&=V_\beta(x,a=1)-V_\beta(x,a=0)\leq 0
\end{aligned}$$

where the above results from the non-decreasing function of $V_\beta(x)$ in x for a given β. Hence, a β-optimal policy is the threshold type and therefore, the cost optimal policy is of the threshold type [35].

Lemma 3: Given the threshold-type policy with the threshold $\overline{X}\in\{1,2,...\}$, then the average cost $C_{avg}(\overline{X},c)$ under the policy is given by

$$C_{avg}(\bar{X},c)=\frac{\omega\left(\frac{\bar{X}^2}{2}+(\frac{1}{p}-\frac{1}{2})\bar{X}+\frac{1}{p^2}-\frac{1}{p}\right)+\frac{c}{p}}{\bar{X}+\frac{1-p}{p}}.$$

Proof: Let $Y(t)$ be the AoI after action in slot t. For $s(t)=x$ and $a(t)=a$, the post-action AoI $Y(t)=x+1-xa\lambda$. The post-action AoI forms a DTMC as shown in Figure 9. We associate each state of the DTMC with a cost. The DTMC incurs a cost of $\omega+c$ in slot t when the post-action AoI in slot t is $Y(t)=1$. It incurs a cost of ωi for states $i\in\{2,3,...,\bar{X}\}$ and $\omega i+c$ for states $i\in\{\bar{X}+1,\bar{X}+2,...\}$. The steady state distribution $\pi=[\pi_1,\pi_2,...]$ of the DTMC is given by

$$\pi_i=\begin{cases}\dfrac{1}{\bar{X}+\frac{1-p}{p}} & \text{if } i=1,2,...,\bar{X};\\[2ex] \dfrac{1}{\bar{X}+\frac{1-p}{p}}(1-p)^{i-\bar{X}} & \text{if } i=\bar{X}+1,...\end{cases}$$

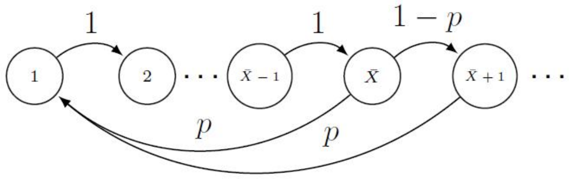

Figure 9. The post-action age $Y(t)$ under the threshold-type policy forms a DTMC.

The average cost of the DTMC is given by

$$C_{avg}(\bar{X},c)=\frac{\omega\left(\frac{\bar{X}^2}{2}+(\frac{1}{p}-\frac{1}{2})\bar{X}+\frac{1}{p^2}-\frac{1}{p}\right)+\frac{c}{p}}{\bar{X}+\frac{1-p}{p}}.$$

Proof of Theorem 4: By definition,

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

$$\text{i.e.,} V_\beta(x,a=0)\ \underset{a=1}{\overset{a=0}{\lessgtr}}\ V_\beta(x,a=1)$$

It follows that

$$x+1+\beta V_\beta(x+1)\ \underset{a=1}{\overset{a=0}{\lessgtr}}\ p(1+c)+(1-p)(x+1+c)\beta pV_\beta(1)+\beta(1-p)V_\beta(x+1)\ ;$$

$$px+\beta pV_\beta(x+1)\ \underset{a=1}{\overset{a=0}{\lessgtr}}\ c+\beta pV_\beta(1)\,.$$

We know that $V_\beta(x)$ is non-decreasing in x for a given β. It follows that if $c \le 0$, the optimal action is to update, i.e., $S(0)=\phi$. For all $c>0$, $\exists x_0$ such that $px+\beta pV_\beta(x+1) > c+\beta pV_\beta(1)$, $\forall x > x_0$. Clearly, the optimal action is to idle till x_0. As c increases, x_0 increases and therefore the set $S(c)$ monotonically increases to the entire state space as $c\to\infty$. It follows that the sub-problem is indexable.

Further, as discussed in the proof outline, we solve

$$C_{avg}(x,c)=C_{avg}(x+1,c)$$

for c to compute the Whittle Index at age-of-information x. This completes the proof.

5.3 *Markovian Channels with CSI*

Proof of Lemma 1:

1. Let f be the stationary deterministic policy of choosing to update the user in every time-slot, if the channel is 'ON'. The state $(X(t),\Lambda(t))$ under this policy f forms a discrete time Markov chain (DTMC) as shown in Figure 10. The steady state distribution $\pi=[\pi_{10},\pi_{11},\pi_{20},\pi_{21},...]$ of the DTMC is given by

$$\pi_{10}=\frac{(1-p)(1-q)}{2-p-q};\pi_{11}=\frac{p(1-q)}{2-p-q};$$

$$\pi_{x0}=q^{x-1}\pi_{10},\forall x\in\{2,3,...\};$$

$$\pi_{x1}=q^{x-2}(1-q)\pi_{10},\forall x\in\{2,3,...\}.$$

The average age 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 result.

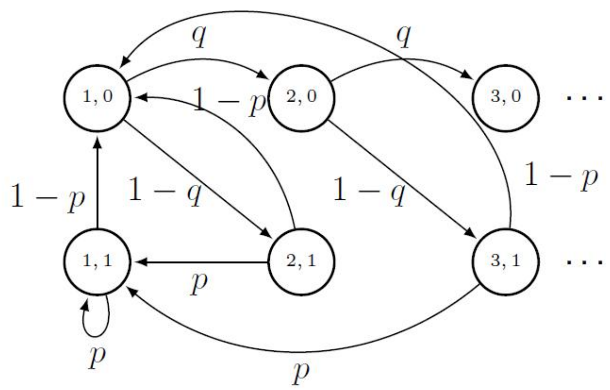

Figure 10. The age $X(t)$ under the policy f forms a DTMC.

2. The value function equation $V_\beta(x,\lambda)$ is given by

$$V_\beta(x,\lambda)=\min\{V_\beta\left((x,\lambda),a=0\right),V_\beta\left((x,\lambda),a=1\right)\};$$

where, $V_\beta\left((x,0),a=0\right)=x+1+\beta\left(qV_\beta(x+1,0)+(1-q)V_\beta(x+1,1)\right)$,

$V_\beta\left((x,0),a=1\right)=x+1+c+\beta\left(qV_\beta(x+1,0)+(1-q)V_\beta(x+1,1)\right)$,

$V_\beta\left((x,1),a=0\right)=x+1+\beta\left(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0)\right)$,

and $V_\beta\left((x,1),a=1\right)=1+c+\beta\left(pV_\beta(1,1)+(1-p)V_\beta(1,0)\right)$.

From above, $V_\beta(x,\lambda)$ is non-decreasing function in age x for a given β and λ. From Proposition 5 of [37] if the first condition of Lemma 1 is satisfied, then for a given state (x,λ) and discount factor β, the quantity $V_\beta(x,\lambda)$ is finite. Therefore, $k=V_\beta(1,0)-V_\beta(1,1)$ is finite. It follows that $V_\beta(x,\lambda)-V_\beta(1,1)\geq k;\forall x,\forall\lambda$. Choose $V_\beta(1,1)$ as a reference state and choose a non-negative L appropriately based on k such that $V_\beta(x,\lambda)-V_\beta(1,1)\geq -L;\forall x,\forall\lambda$.

By verifying the two conditions, the lemma immediately follows.

Proof of Lemma 2: Suppose a β-optimal action for state $(x,1)$ is to update, i.e.,

$$V_\beta\left((x,1),a=1\right)-V_\beta\left((x,1),a=0\right)<0.$$

Then, an β-optimal action for state $(x',1)$, where $x'>x$ is also to update since:

$$\begin{aligned}
&V_\beta\left((x',1),a=1\right)-V_\beta\left((x',1),a=0\right)\\
&=\left(1+c+\beta(pV_\beta(1,1)+(1-p)V_\beta(1,0))\right)-\left(x'+1+\beta(pV_\beta(x'+1,1)+(1-p)V_\beta(x'+1,0))\right)\\
&\leq\left(1+c+\beta(pV_\beta(1,1)+(1-p)V_\beta(1,0))\right)-\left(x+1+\beta(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0))\right)\\
&=V_\beta\left((x,1),a=1\right)-V_\beta\left((x,1),a=0\right)\leq 0,
\end{aligned}$$

where the above results from the non-decreasing function of $V_\beta(x,\lambda)$ in x for a given λ and β. Hence, a β-optimal policy is the threshold type and therefore, the cost optimal policy is of the threshold type [35].

Lemma 4: Given the threshold-type policy with the threshold $\bar{X}\in\{1,2,...\}$, then the average cost $C_{avg}(\bar{X},c)$ under the policy is given by

$$C_{avg}(\bar{X},c)=\left[c+\frac{\bar{X}(\bar{X}+1)}{2}\right](\pi_{10}+\pi_{11})+\left[\frac{\bar{X}}{1-q}+\frac{1}{(1-q)^2}\right]\pi_{\bar{X}0}$$

where, $\pi_{10}=-\dfrac{(p-1)q^2+(p^2-4p+3)q-p^2+3p-2}{(q^2+(p-3)q-p+2)\bar{X}+(p-1)(q+p-1)^{\bar{X}}-p+1}$;

$$\pi_{11}=-\frac{p((1-p)q^2-(p^2-4p+3)q+p^2-3p+2)}{(p-1)((q^2+(p-3)q-p+2)\bar{X}+(q+p-1)^{\bar{X}-1}((p-1)q+p^2-2p+1)-p+1)};$$

$$\pi_{\bar{X}0}=\pi_{10}\left[\frac{p-1}{p+q-2}+\frac{(q-1)(p+q-1)^{\bar{X}-1}}{p+q-2}\right]+\pi_{11}\left[\frac{p-1}{p+q-2}-\frac{(p-1)(p+q-1)^{\bar{X}-1}}{p+q-2}\right].$$

Proof: Let $Y(t)$ be the AoI after action in slot t. For $s(t)=(x(t),\Lambda(t))$ and $a(t)=a$, the post-action AoI $Y(t)=x+1-xa\lambda$. The post-action AoI forms a DTMC as shown in Figure 11. We associate each state of the DTMC with a cost. The DTMC incurs a cost of $\omega+c$ in slot t when the post-action AoI in slot t is $Y(t)=1$. It incurs a cost of ωi for states $\left((i,0),(i,1)\right)$, where as $i\in\{2,3,...\}$.

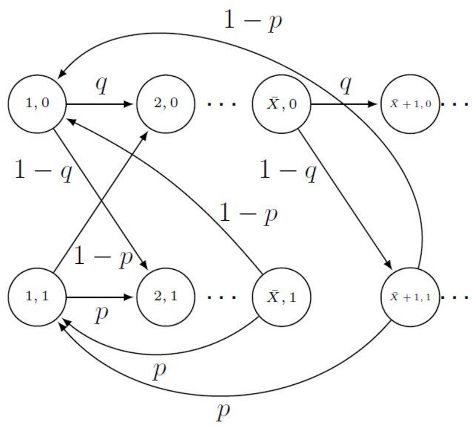

Figure 11. The post-action age $Y(t)$ under the threshold-type policy forms a DTMC.

The steady state distribution $\pi = [\pi_{10}, \pi_{11}, \pi_{20}, \pi_{21}, ...]$ of the DTMC is given by

$$\begin{bmatrix} \pi_{k0} \\ \pi_{k1} \end{bmatrix} = A^{k-1} \begin{bmatrix} \pi_{10} \\ \pi_{11} \end{bmatrix}, k \in \{2, 3, ..., \bar{X}\};$$

where $A = \begin{bmatrix} q & 1-p \\ 1-p & q \end{bmatrix}$ is diagonalizable. We exploit this to explicitly calculate the steady state distribution, i.e., we compute π_{k0}, π_{k1} in terms of $\{\pi_{10}, \pi_{11}\}$ for $k \in \{2, ..., \bar{X}\}$ from the above equation.

Solving the above we get, $\pi_{\bar{X}0}, \pi_{\bar{X}1}$ in terms $\{\pi_{10}, \pi_{11}\}$ of as follows:

$$\pi_{\bar{X}0} = \pi_{10}\left[\frac{p-1}{p+q-2} + \frac{(q-1)(p+q-1)^{\bar{X}-1}}{p+q-2}\right] + \pi_{11}\left[\frac{p-1}{p+q-2} - \frac{(p-1)(p+q-1)^{\bar{X}-1}}{p+q-2}\right]$$

$$\pi_{\bar{X}1} = \pi_{11}\left[\frac{q-1}{p+q-2} + \frac{(p-1)(p+q-1)^{\bar{X}-1}}{p+q-2}\right] + \pi_{10}\left[\frac{q-1}{p+q-2} - \frac{(q-1)(p+q-1)^{\bar{X}-1}}{p+q-2}\right]$$

$$\pi_{\bar{X}+k0} = q^k \pi_{\bar{X}0}, \forall k \in \{1, 2, ...\};$$

$$\pi_{\bar{X}+k1} = q^{k-1}(1-q)\pi_{\bar{X}0}, \forall k \in \{1, 2, ...\};$$

As $\pi_{\bar{X}0}$, we can write in terms of $\{\pi_{10}, \pi_{11}\}$ from above, which makes every state in the Markov chain be represented a function of $\{\pi_{10}, \pi_{11}\}$. Furthermore, solving the below two equations, we can obtain the stationary distribution

$$\pi_{10} = (1-p)\sum_{i=\bar{X}}^{\infty} \pi_{i1}; \pi_{11} = p\sum_{i=\bar{X}}^{\infty} \pi_{i1}$$

$$\sum_{i=1}^{\infty} (\pi_{i0} + \pi_{i1}) = 1.$$

Solving the two equations π_{10}, π_{11} obtained as follows:

$$\pi_{10} = -\frac{(p-1)q^2 + (p^2 - 4p + 3)q - p^2 + 3p - 2}{(q^2 + (p-3)q - p + 2)\bar{X} + (p-1)(q+p-1)^{\bar{X}} - p + 1};$$

$$\pi_{11} = -\frac{p((1-p)q^2 - (p^2 - 4p + 3)q + p^2 - 3p + 2)}{(p-1)((q^2 + (p-3)q - p + 2)\bar{X} + (q+p-1)^{\bar{X}-1}((p-1)q + p^2 - 2p + 1) - p + 1)};$$

Substitute π_{10}, π_{11} and $\pi_{\bar{X}0}$ values in the below equation to obtain the final result.

$$C_{avg}(\bar{X},c)=\left[c+\frac{\bar{X}(\bar{X}+1)}{2}\right](\pi_{10}+\pi_{11})+\left[\frac{\bar{X}}{1-q}+\frac{1}{(1-q)^2}\right]\pi_{\bar{X}0}$$

Proof of Theorem 5:

Let us define $S(c)=S_1(c)\cup S_2(c),$ where $S_1(c)$ be the set of $(x,0)$ for which the optimal action is idle. Similarly, let $S_2(c)$ be the set of states of the form $(x,1),$ for which the optimal action is idle. Recall that,

$$V_\beta\left((x,0),a=0\right)\ \overset{a=0}{\underset{a=1}{\lessgtr}}\ V_\beta\left((x,0),a=1\right),\text{i.e.,}$$

$$\left(x+1+\beta(qV_\beta(x+1,0)+(1-q)V_\beta(x+1,1))\right)$$
$$\overset{a=0}{\underset{a=1}{\lessgtr}}\left(x+1+c+\beta(qV_\beta(x+1,0)+(1-q)V_\beta(x+1,1))\right).$$

For state $(x,0)$ the optimal action is to idle for $c\geq 0.$ So corresponding set $S_1(c)=\{(x,0):x=1,2,...\}.$ Now for state $(x,1)$,

$$V_\beta\left((x,1),a=0\right)\ \overset{a=0}{\underset{a=1}{\lessgtr}}\ V_\beta\left((x,1),a=1\right)$$
$$\Rightarrow\left(x+1+\beta(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0))\right)$$
$$\overset{a=0}{\underset{a=1}{\lessgtr}}\left(1+c+\beta(pV_\beta(1,1)+(1-p)V_\beta(1,0))\right).$$

We know that $V_\beta(x,\lambda)$ is non-decreasing in x for a given λ , β. For $c\leq 0$, optimal action is to update, i.e., $S(0)=\phi.$ Note that $\forall c>0,\exists x_0$ such that $x+\beta(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0))>c+\beta(pV_\beta(1,1)+(1-p)\ \ V_\beta(1,0)),\ \forall x>x_0.$ Clearly, the optimal action is to idle till state x_0. As c increases, x_0 increases, so the set $S_2(c)=\{(x,1):x=1,2,...x_0\}$ monotonically increases. Hence, the set $S(c)$ monotonically increases to the entire state space as $c\to\infty$ and the sub-problem is indexable.

The Whittle Index for state $(x,0)$ is $I(x,0)=0,$ as both the actions result same values for $C=0.$ For states of the form $(x,1),$ as discussed in the proof outline, we solve $C_{avg}(x,c)=C_{avg}(x+1,c),$ for c to compute the Whittle Index at age-of-information $x.$ This completes the proof.

5.4 *i.i.d. Channels with CSI*

It is a special case of Markovian channels with CSI and the proof details in this setting are similar to that of i.i.d. channels without CSI. We skip the details due to lack of space.

5.5 *Markovian Channels with Delayed-CSI*

Proof of Lemma 1: 1) Let f be the stationary deterministic policy of always choosing to update. The state $(X(t),\Lambda(t-D))$ under this policy f forms a discrete time Markov chain (DTMC) as shown in Figure 12 (For each transition probability omitted $(D-1)$ in figure for better readability). The steady state distribution $\pi=[\pi_{10},\pi_{11},\pi_{20},\pi_{21},...]$ of the DTMC is given by

$$\begin{bmatrix} \pi_{k0} \\ \pi_{k1} \end{bmatrix} = A^{k-1} \begin{bmatrix} \pi_{10} \\ \pi_{11} \end{bmatrix}, k \in \{2,3,...\};$$

where $A = \begin{bmatrix} q\mathrm{P}_{00}(D-1) & (1-p)\mathrm{P}_{00}(D-1) \\ (1-q)\mathrm{P}_{10}(D-1) & p\mathrm{P}_{10}(D-1) \end{bmatrix}$ is diagonalizable. We exploit this to explicitly calculate the steady state distribution, i.e., we compute π_{k0}, π_{k1} in terms of $\{\pi_{10}, \pi_{11}\}$ for $k \in \{2,3,...\}$ from the above equation. Now you will have everything in terms of $\{\pi_{10}, \pi_{11}\}$. Substituting this you can write π_{10} in terms of π_{11} and the sum of stationary probabilities equal to 1, and solve these two equations to obtain steady state distribution. Using the steady state distribution, we can calculate the average age which is given as follows

$$\sum_{i=1}^{\infty} i(\pi_{i0} + \pi_{i1})$$

We observe that $\pi_{(k+1)0} + \pi_{(k+1)1} < \pi_{k0} + \pi_{k1}; \forall k \in \mathrm{Z}^+$. Further, as age i increases, $(\pi_{i0} + \pi_{i1})$ decreases exponentially. It therefore follows that the summation is bounded.

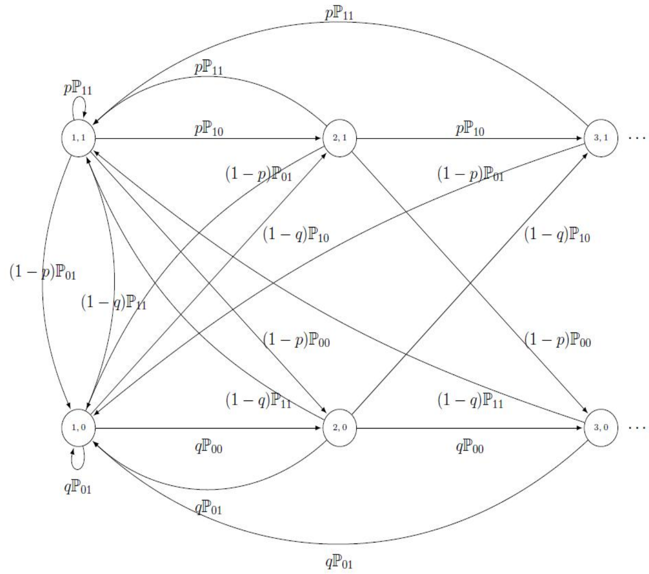

Figure 12. The age $X(t)$ under the policy f forms a DTMC.

On the other hand, the average updating cost is c. Hence, the average cost per time slot under this policy is bounded, thus proving the desired result.

From Proposition 5 of [37] for a given state (x, λ) and discount factor β, the quantity $V_\beta(x, \lambda)$ is finite.

$V_\beta(x,\lambda) = \min\{V_\beta\left((x,\lambda), a=0\right), V_\beta\left((x,\lambda), a=1\right)\};$

where, $V_\beta\left((x,0), a=0\right) = x+1+\beta\left(qV_\beta(x+1,0)+(1-q)V_\beta(x+1,1)\right),$

$$V_\beta\left((x,1), a=0\right) = x+1+\beta\left(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0)\right),$$

$$V_\beta\left((x,0), a=1\right) = c+q\mathrm{P}_{01}(D-1)+(1-q)\mathrm{P}_{11}(D-1)+(x+1)(q\mathrm{P}_{00}(D-1)+(1-q)\mathrm{P}_{10}(D-1))+\beta[p\mathrm{P}_{11}(D-1)V_\beta(1,1) \\ +(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)+p\mathrm{P}_{10}(D-1)V_\beta(x+1,1)+(1-p)\mathrm{P}_{00}(D-1)V_\beta(x+1,0)],$$

$$V_\beta\left((x,1), a=1\right) = c+p\mathrm{P}_{11}(D-1)+(1-p)\mathrm{P}_{01}(D-1)+(x+1)(p\mathrm{P}_{10}(D-1)+(1-p)\mathrm{P}_{00}(D-1))+\beta[p\mathrm{P}_{11}(D-1)V_\beta(1,1) \\ +(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)+p\mathrm{P}_{10}(D-1)V_\beta(x+1,1)+(1-p)\mathrm{P}_{00}(D-1)V_\beta(x+1,0)].$$

From above $V_\beta(x,\lambda)$ is non-decreasing function in AoI x for a given β and λ. From Proposition 5 of [37], if the first condition of Lemma 1 is satisfied, then for a given state (x,λ) and discount factor β, the quantity $V_\beta(x,\lambda)$ is finite. Therefore, $k = V_\beta(1,0) - V_\beta(1,1)$. It follows that $V_\beta(x,\lambda) - V_\beta(1,1) \geq k; \forall x, \forall \lambda$. We choose $V_\beta(1,1)$ as a reference state and choose a non-negative L appropriately based on k to prove the result.

By verifying the two conditions, the lemma immediately follows.

Proof of Lemma 2: Suppose a β -optimal action for state $(x,1)$ is to update, i.e.,

$$V_\beta\left((x,1),a=1\right)-V_\beta\left((x,1),a=0\right)<0.$$

Then, an β -optimal action for state $(x',1)$, where $x'>x$ is also to update since:

$$\begin{aligned}
&V_\beta\left((x',1),a=1\right)-V_\beta\left((x',1),a=0\right)\\
&=c+p\mathrm{P}_{11}(D-1)+(1-p)\mathrm{P}_{01}(D-1)+(x'+1)(p\mathrm{P}_{10}(D-1)+(1-p)\mathrm{P}_{00}(D-1)-1)+\beta[p\mathrm{P}_{11}(D-1)V_\beta(1,1)\\
&\quad+(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)+p(\mathrm{P}_{10}(D-1)-1)V_\beta(x'+1,1)+(1-p)(\mathrm{P}_{00}(D-1)-1)V_\beta(x'+1,0)]\\
&\le c+p\mathrm{P}_{11}(D-1)+(1-p)\mathrm{P}_{01}(D-1)+(x+1)(p\mathrm{P}_{10}(D-1)+(1-p)\mathrm{P}_{00}(D-1)-1)+\beta[p\mathrm{P}_{11}(D-1)V_\beta(1,1)\\
&\quad+(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)+p(\mathrm{P}_{10}(D-1)-1)V_\beta(x+1,1)+(1-p)(\mathrm{P}_{00}(D-1)-1)V_\beta(x+1,0)]\\
&=V_\beta\left((x,1),a=1\right)-V_\beta\left((x,1),a=0\right)\le 0
\end{aligned}$$

where the above results from the non-decreasing function of $V_\beta(x,\lambda)$ in x for a given λ and β. Similarly we can show that if β -optimal action for state $(x,0)$ is to update, then β -optimal action for state $(x',0)$ is also to update for $x'>x$. Hence, a β -optimal policy is the threshold type and therefore, the cost optimal policy is of the threshold type [35].

Proof of Theorem 8:

$$V_\beta(x,1)=\min\{V_\beta\left((x,1),a=0\right),V_\beta\left((x,1),a=1\right)\};$$

$$\begin{aligned}
&x+1+\beta\left(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0)\right)\\
&\mathop{\gtrless}_{a=1}^{a=0}\; c+p\mathrm{P}_{11}(D-1)+(1-p)\mathrm{P}_{01}(D-1)+(x+1)(p\mathrm{P}_{10}(D-1)+(1-p)\mathrm{P}_{00}(D-1))+\beta\big[p\mathrm{P}_{11}(D-1)V_\beta(1,1)\\
&\quad+(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)+p\mathrm{P}_{10}(D-1)V_\beta(x+1,1)+(1-p)\mathrm{P}_{00}(D-1)V_\beta(x+1,0)\big]\\
&\Rightarrow (x+1)\left[1-p\mathrm{P}_{10}(D-1)-(1-p)\mathrm{P}_{00}(D-1)\right]+\beta\left[p(1-\mathrm{P}_{10}(D-1))V_\beta(x+1,1)+(1-p)(1-\mathrm{P}_{01}(D-1))V_\beta(x+1,0)\right]\\
&\mathop{\gtrless}_{a=1}^{a=0}\; c+p\mathrm{P}_{11}(D-1)+(1-p)\mathrm{P}_{01}(D-1)+\beta\left[p\mathrm{P}_{11}(D-1)V_\beta(1,1)+(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)\right];
\end{aligned}$$

where $\mathrm{P}_{ij}(D-1)=\mathrm{P}(\Lambda(t)=j\,|\Lambda(t-D+1)=i)$, for $i\in\{0,1\}$, $j\in\{0,1\}$.

We know that $V_\beta(x,1)$ is non-decreasing in x for a given β. For $c\le 0$, the optimal action is to update, i.e., $S(0)=\phi$. Note that $\forall c>0,\ \exists x_0$ such that $(x+1)\left[1-p\mathrm{P}_{10}(D-1)-(1-p)\mathrm{P}_{00}(D-1)\right]$ $+\beta\left[p(1-\mathrm{P}_{10}(D-1))V_\beta(x+1,1)+(1-p)(1-\mathrm{P}_{01}(D-1))V_\beta(x+1,0)\right]$ $>c+p\mathrm{P}_{11}(D-1)+(1-p)\mathrm{P}_{01}(D-1)+\beta\left[p\mathrm{P}_{11}(D-1)V_\beta(1,1)+(1-p)\mathrm{P}_{01}(D-1)V_\beta(1,0)\right];$ $\forall x>x_0$ i.e., the optimal action is to idle till x_0. As c increases, x_0 increases, so the set $S_2(c)=\{(x,1):x=1,2,\ldots,x_0\}$ monotonically increases. Similarly for states of the form $(x,0)$, as c increases, the set $S_1(c)=\{s=(x,0):x=1,2,\ldots\}$ monotonically increases. The union of both these disjoint sets monotonically increases and therefore, the set $S(c)$ monotonically increases to the entire state space as $c\to\infty$.

Funding

This work is supported by a SERB grant on Versatile Scheduling Policies with Provable Guarantees for Latency Sensitive Systems.

Conflict of Interest

There is no conflict of interest for this study.

References

[1] Kaul, S.; Gruteser, M.; Rai, V.; Kenney, J. Minimizing age of information in vehicular networks. In Proceedings of 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, Salt Lake City, UT, USA, 27–30 June 2011. https://doi.org/10.1109/SAHCN.2011.5984917.

[2] He, Q.; Yuan, D.; Ephremides, A. Optimizing freshness of information: On minimum age link scheduling in wireless systems. In Proceedings of 2016 14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Tempe, AZ, USA, 9–13 May 2016. https://doi.org/10.1109/WIOPT.2016.7492912.

[3] Talak, R.; Karaman, S.; Modiano, E. Optimizing information freshness in wireless networks under general interference constraints. In Proceedings of Mobihoc '18: The Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing, Los Angeles, CA, USA, 26–29 June 2018. https://doi.org/10.1145/3209582.3209589.

[4] Jhunjhunwala, P.R.; Sombabu, B.; Moharir, S. Optimal AoI-aware scheduling and cycles in graphs. *IEEE Trans Commun.* 2020, *68*, 1593–1603, https://doi.org/10.1109/TCOMM.2019.2962038.

[5] Tripathi, V.; Modiano, E. A Whittle Index Approach to Minimizing Functions of Age of Information. In Proceedings of 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 24–27 September 2019. https://doi.org/10.1109/ALLERTON.2019.8919842.

[6] Sombabu, B.; Moharir, S. Age-of-Information Based Scheduling for Multi-Channel Systems. *IEEE Trans. Wirel. Commun.* 2020, *19*, 4439–4448, https://doi.org/10.1109/twc.2020.2983670.

[7] Tripathi, V.; Moharir, S. Age of information in multi-source systems. In Proceedings of GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017. https://doi.org/10.1109/GLOCOM.2017.8254578.

[8] Hsu, Y.-P.; Modiano, E.; Duan, L. Scheduling Algorithms for Minimizing Age of Information in Wireless Broadcast Networks with Random Arrivals. *IEEE Trans. Mob. Comput.* 2019, *19*, 2903–2915, https://doi.org/10.1109/tmc.2019.2936199.

[9] Kadota, I.; Sinha, A.; Modiano, E. Optimizing Age of Information in Wireless Networks with Throughput Constraints. In Proceedings of IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, Honolulu, HI, USA, 16–19 April 2018. https://doi.org/10.1109/INFOCOM.2018.8486307.

[10] Gao, W. Cao, G. Srivatsa, M. Iyengar, A. Distributed maintenance of cache freshness in opportunistic mobile networks. In Proceedings of 2012 IEEE 32nd International Conference on Distributed Computing Systems, Macau, China, 18–21 June 2012. https://doi.org/10.1109/ICDCS.2012.63.

[11] Yates, R.D.; Ciblat, P.; Yener, A.; Wigger, M. Age-optimal constrained cache updating. In Proceedings of 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017. https://doi.org/10.1109/ISIT.2017.8006506.

[12] Niyato, D.; Kim, D.I.; Wang, P.; Song, L. A novel caching mechanism for Internet of Things (IoT) sensing service with energy harvesting. In Proceedings of 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016. https://doi.org/10.1109/ICC.2016.7511440.

[13] Yates, R.D. Lazy is timely: Status updates by an energy harvesting source. In Proceedings of 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015. https://doi.org/10.1109/ISIT.2015.7283009.

[14] Arafa, A.; Yang, J.; Ulukus, S.; Poor, H.V. Age-Minimal Transmission for Energy Harvesting Sensors with Finite Batteries: Online Policies. *IEEE Trans. Inf. Theory* 2019, *66*, 534–556, https://doi.org/10.1109/tit.2019.2938969.

[15] Bacinoglu, B.T.; Ceran, E.T.; Uysal-Biyikoglu, E. Age of information under energy replenishment constraints. In Proceedings of 2015 Information Theory and Applications Workshop (ITA), San Diego, CA, USA, 1–6 February 2015. https://doi.org/10.1109/ITA.2015.7308962.

[16] Costa, M.; Valentin, S.; Ephremides, A. On the age of channel information for a Finite-State Markov model. In Proceedings of 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015. https://doi.org/10.1109/ICC.2015.7248966.

[17] Farazi, S.; Klein, A.G.; Brown, D.R. On the average staleness of global channel state information in wireless networks with random transmit node selection. In Proceedings of 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 20–25 March 2016. https://doi.org/10.1109/ICASSP.2016.7472352.

[18] Talak, R.; Karaman, S.; Modiano, E. Optimizing age of information in wireless networks with perfect channel state information. In Proceedings of 2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Shanghai, China, 7–11 May 2018. https://doi.org/10.23919/WIOPT.2018.8362818.

[19] Sombabu, B.; Mate, A.; Manjunath, D.; Moharir, S. Whittle index for AoI-aware scheduling. In Proceedings of 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), Bengaluru, India, 7–11 January 2020. https://doi.org/10.1109/COMSNETS48256.2020.9027444.

[20] Papadimitriou, C.H.; Tsitsiklis, J.N. The Complexity of Optimal Queuing Network Control. *Math. Oper. Res.* 1999, *24*, 293–305, https://doi.org/10.1287/moor.24.2.293.

[21] Whittle, P. Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* 1988, *25*, 287–298, https://doi.org/10.2307/3214163.

[22] Chen, M.; Wu, K.; Song, L. A Whittle Index Approach to Minimizing Age of Multi-Packet Information in IoT Network. *IEEE Access* 2021, *9*, 31467–31480, https://doi.org/10.1109/access.2021.3059966.

[23] Tang, H.; Wang, J.; Tang, Z.; Song, J. Scheduling to minimize age of synchronization in wireless broadcast networks with random updates. *IEEE Trans. Wirel. Commun.* 2020, *19*, 4023–4037, https://doi.org/10.1109/TWC.2020.2979436.

[24] Chen, Y.; Tang, H.; Wang, J.; Song, J. Optimizing Age Penalty in Time-Varying Networks with Markovian and Error-Prone Channel State. *Entropy* 2021, *23*, 91, https://doi.org/10.3390/e23010091.

[25] Hatami, M.; Leinonen, M.; Chen, Z.; Pappas, N.; Codreanu, M. On-Demand AoI Minimization in Resource-Constrained Cache-Enabled IoT Networks with Energy Harvesting Sensors. *IEEE Trans. Commun.* 2022, *70*, 7446–7463, https://doi.org/10.1109/tcomm.2022.3208873.

[26] Tang, H.; Wang, J.; Song, L.; Song, J. Minimizing Age of Information with Power Constraints: Multi-User Opportunistic Scheduling in Multi-State Time-Varying Channels. *IEEE J. Sel. Areas Commun.* 2020, *38*, 854–868, https://doi.org/10.1109/jsac.2020.2980911.

[27] Buyukates, B.; Ulukus, S. Age of information with Gilbert-Elliot servers and samplers. In Proceedings of 2020 54th Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, USA, 18–20 March 2020. https://doi.org/10.1109/CISS48834.2020.1570627498.

[28] Sombabu, B.; Moharir, S. Age-of-information aware scheduling under Markovian energy arrivals. In Proceedings of 2020 International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, 19–24 July 2020. https://doi.org/10.1109/SPCOM50965.2020.9179542.

[29] Aggarwal, S.; uz Zaman, M.A.; Bastopcu, M.; Başar, T. Weighted age of information based scheduling for l arge population games on networks. arXiv preprint arXiv:2209.12888, https://doi.org/10.48550/arXiv.2209.12888.

[30] Kosta, A.; Pappas, N.; Ephremides, A.; Angelakis, V. Age and value of information: Non-linear age case. I n Proceedings of 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017. https://doi.org/10.1109/ISIT.2017.8006543.

[31] Sun, J.; Jiang, Z.; Krishnamachari, B.; Zhou, S.; Niu, Z. Closed-Form Whittle's Index-Enabled Random Ac cess for Timely Status Update. *IEEE Trans. Commun.* 2019, *68*, 1538–1551, https://doi.org/10.1109/tcomm.2019.2960346.

[32] Sun, J.; Jiang, Z.; Zhou, S.; Niu, Z. Optimizing Information Freshness in Broadcast Network with Unreliable Links and Random Arrivals: An Approximate Index Policy. In Proceedings of IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 29 April–2 May 2019. https://doi.org/10.1109/INFCOMW.2019.8845145.

[33] Maatouk, A.; Kriouile, S.; Assaad, M.; Ephremides, A. On the Optimality of the Whittle's Index Policy for Minimizing the Age of Information. *IEEE Trans. Wirel. Commun.* 2020, *20*, 1263–1277, https://doi.org/10.1109/twc.2020.3032237.

[34] Kadota, I.; Sinha, A.; Uysal-Biyikoglu, E.; Singh, R.; Modiano, E. Scheduling Policies for Minimizing Age of Information in Broadcast Wireless Networks. *IEEE/ACM Trans. Netw.* 2018, *26*, 2637–2650, https://doi.org/10.1109/tnet.2018.2873606.
[35] Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 1994, pp. 392–394.
[36] Whittle, P. Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* 1988, *25*, 287–298, https://doi.org/10.2307/3214163.
[37] Sennott, L.I. Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes with Unbounded Costs. *Oper. Res.* 1989, *37*, 626–633, https://doi.org/10.1287/opre.37.4.626.

Appendix A: Efficient Scheduling for I.I.D. Channels without CSI

The scheduling algorithm for systems without CSI involves calculating the Whittle Index for every sensor in a time slot. Tracking every sensor becomes a bottleneck in employing the index based algorithm for large scale sensor systems. We show efficient implementations of the vanilla Whittle Index algorithm that don't require calculating the index of every sensor and incur little to no additional age. Note that the Whittle Index for the given setting is monotone increasing in the age x, channel ON probability p and weight ω and hence this structure can be used to decide the subset of total sensors that need to be tracked at every time slot.

Definition 10: (Standby time) The standby time t_i of a sensor i after it successfully updates the monitoring station is the total time before it's Whittle Index becomes the highest.

Definition 11: (Rank) A sensor i is assigned a rank R_i higher than the rank R_j of a sensor j $(R_i < R_j)$ if $\omega_i > \omega_j$ and $p_i > p_j$. Otherwise $R_i = R_j$.

Lemma 5: (Lower bound on standby time) The standby time t_i of a sensor i having a rank R_i is lower bounded by $|S_i|$ where $S_i = \{j : R_j < R_i\}$.

Proof: The Whittle Index $I(x_i, \omega_i, p_i)$ is monotone increasing in x_i, ω_i and p_i. If a sensor i makes a successful update, it's age x_i will be less than the age of a sensor j until it is scheduled i.e. $x_i < x_j$ until sensor j is scheduled successfully. By the definition of rank, $\omega_i > \omega_j$ and $p_i > p_j$ if $R_i = R_j$. Hence using the monotonicity of $I(x_i, \omega_i, p_i)$ it is obvious that for the most recently updated sensor i, $I(x_i, \omega_i, p_i) < I(x_j, \omega_j, p_j)$ for every j such that $R_j < R_i$. In other words the Whittle Index of a sensor that transmits an update cannot equate the Whittle Index of higher ranked sensors until they are scheduled successfully. Since there are $|S_i|$ higher ranked sensors the lower bound of standby time can be set to be $|S_i|$.

Hence Lemma 5 indicates the minimum time it takes for a sensor to achieve the highest Whittle Index after it is successfully scheduled. Next we characterize the time it takes for the sensor with the most recent update to equate the Whittle Index of lower ranked sensors.

Lemma 6: Let sensor i successfully transmit an update in time slot t. The time taken for sensor i to equate the Whittle Index of a sensor j having age x_j at t such that $R_i < R_j$ and assuming that sensor j remains unscheduled is at least $t_{i,j}$ where, $t_{i,j} = 1 + \dfrac{2(\omega_j - \omega_i + p_j \omega_j x_j)}{p_i \omega_i - p_j \omega_j}$.

Proof: Let $T_{i,j}$ be the number of time-slots a successfully scheduled sensor i takes to equate the Whittle Index of a lower ranked sensor j. So Whittle Index of i, j after y unscheduled time-slots will be $\omega_i \left(\dfrac{p_i y^2}{2} - \dfrac{p_i y}{2} + y \right)$, $\omega_j \left(\dfrac{p_j (x_j + y)^2}{2} - \dfrac{p_j (x_j + y)}{2} + x_j + y \right)$ respectively. Equate them and solve the quadratic in y to obtain a positive solution $T_{i,j}$, which is always greater than $t_{i,j}$ as defined in Lemma 6.

Lemma 6 thus helps in characterizing the relative order of the Whittle Index of the sensors without explicitly calculating their index. Note that Lemma 6 assumes that the lower ranked sensors remain unscheduled. However that may not hold true and hence we impose an upper bound on the standby time to account for this assumption. The upper bound on standby time for a sensor that successfully transmits an update is then a measure of the total time it takes for all other sensors to be successfully transmitted.

We propose the first implementation of the vanilla Whittle Index policy in Algorithm 1 which we call the Efficient Conservative scheduling. Here we set an upper bound of $n-1$ on the standby time where n denotes the total number of sensors. We call it conservative as $n-1$ is the minimum time it takes to schedule $n-1$ sensors and thus the upper bound is tight. Note that in Algorithm 1 we only compute the Whittle Index of the active set in every time slot.

Algorithm 1 Efficient Conservative Scheduling

Input: Number of sensors $n, p_i, \omega_i, R_i; 1 \le i \le n, T$
Initialize: Active $\leftarrow [1, \ldots, n]$
Standby $\leftarrow [\ ]$ {heap of tuple (Timestamp, sensor index)}
For each t in $1 \ldots T$:
    If $t ==$ Standby[0][0]
        $S ==$ Standby[0][1]
        Standby.pop(), Active.add(S)
        $i =$ Sensor having highest White index in Active
    If i successfully scheduled:
        equate_time $= [t_{i,j}\ \forall j\ s.t.\ R_i < R_j]$
        $t_i = \min(\max(\max(\text{equate\_time}), |S_i|), n-1)$
        $\text{Timestamp}_i = t + t_i$
        Standby.add((Timestamp$_i$, i))

Lemma 7: Let $\{X_i\}_{i=1}^{n}$ be set of independent geometric random variables with mean $\frac{1}{p_i}$ respectively. Let $p_{\min} = \min_{1 \le i \le n} p_i$, $p_{\max} = \max_{1 \le i \le n} p_i$, $X = \sum_{i=1}^{n} X_i$ and $\mu_n = \sum_{i=1}^{n} \mathrm{E}[X_i]$.

Then,

$$\mathrm{P}(X \le \in \mu_n) \le e^{-n\mu_n\left(\frac{p_{\min}}{p_{\max}}\right)(\in -1-\log \in)}$$

Proof:

$$\mathrm{E}[e^{-sX_i}] = \frac{p_i}{e^s - 1 + p_i} \le \left(1 + (s/p_i)\right)^{-1}, \forall s > 0$$

$$\Rightarrow \mathrm{E}[e^{-sX}] \le \prod_{i=1}^{n} \left(1 + (s/p_i)\right)^{-1}$$

Using Markov's inequality,

$$\begin{aligned}\mathrm{P}(X \le \in \mu_n) &\le e^{s \in \mu_n} \mathrm{E}[e^{-sX}] \\ &= \exp\left(s \in \mu_n - \sum_{i=1}^{n} \log(1 + (s/p_i))\right).\end{aligned}$$

Note that $x \log(1 + (s/x))$ is an increasing function in x for $x > 0, s > 0$. Therefore, $p_i \log\left(1 + (s/p_i)\right) \ge p_{\min} \log\left(1 + (s/p_{\min})\right)$.

$$\begin{aligned}\mathrm{P}(X \le \in \mu_n) &\le e^{\left(s \in \mu_n - p_{min} \sum_{i=1}^{n} (1/p_i) \log(1 + (s/p_{min}))\right)} \\ &= \exp\left(s \in \mu_n - p_{min}\mu_n \log(1 + (s/p_{min}))\right).\end{aligned}$$

Setting $s = (\in^{-1} - 1)p_{min}$ yields, $\mathrm{P}(X \le \in \mu_n) \le e^{-p_{min}\mu_n(\in -1-\log \in)} \le e^{-n\mu_n\left(\frac{p_{min}}{p_{max}}\right)(\in -1-\log \in)}$.

We impose a loser upper bound in the next implementation by assuming that the unscheduled sensors follow a first come first serve (FCFS) queue. In our next result we derive probably approximately correct (PAC) scheduling times for such queues.

Lemma 8: Let n be the set of sensors. Let $p_{\min}^n$ and $p_{\max}^n$ denote the minimum and maximum channel ON probability of the sensors respectively. Let δ be the desired prediction error probability. Then assuming that the sensors in n follow a FCFS queue, the total scheduling time X for the queue can be predicted as

$$\mathrm{P}(X \geq \mu_n \in_n^*) \geq 1-\delta,$$

where $\mu_n = \sum_{i=1}^{n} \frac{1}{p_i}, \in_n^* = -W\left(-\exp\left(-\frac{p_{\max}^n \log(\frac{1}{\delta})}{p_{\min}^n |n|} - 1\right)\right)$. Here $W(x)$ is the Lambert function.

Proof: Substitute $\delta = e^{-n\mu_n(\frac{p_{min}}{p_{max}})(\in - 1 - \log \in)}$ from Lemma 7 and solve for $\in$ to get the result.

In the next implementation we upper bound the standby times of a successfully scheduled sensor i with $U_i = \mu_{n/i} \in_{n/i}^*$. Where, n/i denotes the set of all sensors except i. We call this Efficient PAC scheduling and describe it in Algorithm 2.

Algorithm 2 Efficient PAC Scheduling

1: $Input$: Number of sensors $n, p_i, \omega_i, R_i; 1 \leq i \leq n, T$
2: $Initialize$: Active $\leftarrow [1,...,n]$
3: Standby $\leftarrow$ [ ] {heap of tuple (Timestamp, sensor index)}
4: For each t in $1...T$:
5: If $t ==$ Standby[0][0]
6: $S ==$ Standby[0][1]
7: Standby.pop(), Active.add(S)
8: $i =$ Sensor having highest White index in Active
9: If i successfully scheduled:
10: equate_time $= [t_{i,j} \;\; \forall j \;\; s.t. \;\; R_i < R_j]$
11: $t_i = \min(\max(\max(\text{equate\_time}), |S_i|), U_i)$
12: $\text{Timestamp}_i = t + t_i$
13: Standby.add((Timestamp$_i$, i))

In the next two figures we compare the vanilla implementation of the Whittle Index policy with the two efficient implementations we design through simulations. We vary the sensors from 100 to 250 and schedule them for $T = 10000$ time slots. The weights of the sensors are chosen uniformly between 0 and 1 and are normalized so they add to unity. Similarly the channel ON probabilities are also chosen uniformly from 0.3 to 0.9. For the Efficient PAC implementation we set $\delta = 0.05$. In Figure 13, we compare the average cost incurred by the different implementations.

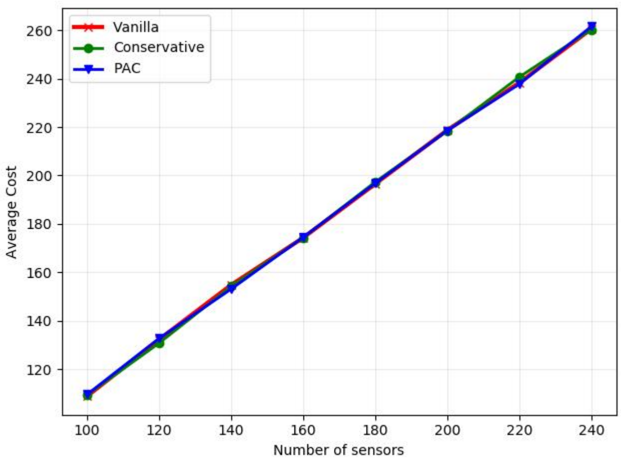

Figure 13. Average cost of different implementations.

All the implementations are observed to have roughly equal cost for all systems. Next we compare the time taken by each implementation in Figure 14.

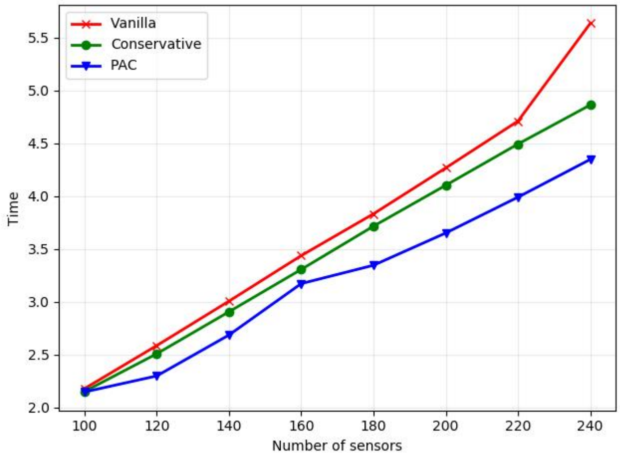

Figure 14. Time taken by different implementations.

It is evident that both the Efficient Conservative and Efficient PAC implementation have better time efficiency than the Vanilla implementation. Moreover the PAC implementation surpasses all the other implementations and it's performance improves with the size of the system.

Appendix B: Markovain Channels with One Delayed-CSI

Proof of Lemma 1: 1) Let f be the stationary deterministic policy of always choosing to update. The state $(X(t), \Lambda(t))$ under this policy f forms a discrete time Markov chain (DTMC) as shown in Figure 15. The steady state distribution $\pi = [\pi_{11}, \pi_{20}, \pi_{30}, \pi_{40}, ...]$ of the DTMC is given by

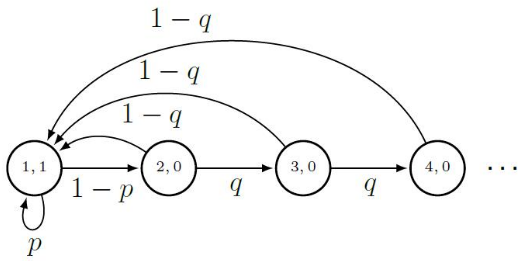

Figure 15. The age $X(t)$ under the policy f forms a DTMC.

The average age is given by

$$\pi_{11} + \sum_{i=1}^{\infty} i\pi_{i0} = 1 + \frac{1-p}{(1-q)(2-p-q)}$$

On the other hand, the average updating cost is c. Hence, the average cost per time slot under this policy is

$$1 + c + \frac{1-p}{(1-q)(2-p-q)} < \infty$$

which is finite and yields the result.

2) From Proposition 5 of [37] for a given state (x, λ) and discount factor β, the quantity $V_\beta(x, \lambda)$ is finite.

$$V_\beta(x, \lambda) = \min\{V_\beta\left((x, \lambda), a = 0\right), V_\beta\left((x, \lambda), a = 1\right)\};$$

where, $V_\beta\left((x, 0), a = 0\right) = x + 1 + \beta\left(qV_\beta(x+1, 0) + (1-q)V_\beta(x+1, 1)\right),$

$$V_\beta\left((x,0),a=1\right)=qx+1+c+\beta\left(qV_\beta(x+1,0)+(1-q)V_\beta(1,1)\right),$$

$$V_\beta\left((x,1),a=0\right)=x+1+\beta\left(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0)\right),$$

and $V_\beta\left((x,1),a=1\right)=(1-p)x+1+c+\beta\left(pV_\beta(1,1)+(1-p)V_\beta(x+1,0)\right).$

From above $V_\beta(x,\lambda)$ is non-decreasing function in AoI x for a given β and λ. From Proposition 5 of [37], if the first condition of Lemma 1 is satisfied, then for a given state (x,λ) and discount factor β, the quantity $V_\beta(x,\lambda)$ is finite. Therefore, $k=V_\beta(1,0)-V_\beta(1,1)$. It follows that $V_\beta(x,\lambda)-V_\beta(1,1)\geq k;\forall x,\forall\lambda$. We choose $V_\beta(1,1)$ as a reference state and choose a non-negative L appropriately based on k to prove the result.

By verifying the two conditions, the lemma immediately follows.

Proof of Lemma 2: Suppose a β-optimal action for state $(x,1)$ is to update, i.e.,

$$V_\beta\left((x,1),a=1\right)-V_\beta\left((x,1),a=0\right)<0.$$

Then, an β-optimal action for state $(x',1)$, where $x'>x$ is also to update since:

$$\begin{aligned}
&V_\beta\left((x',1),a=1\right)-V_\beta\left((x',1),a=0\right)\\
&=\left((1-p)x'+1+c+\beta(pV_\beta(1,1)+(1-p)V_\beta(x'+1,0))\right)\\
&\quad-\left(x'+1+\beta(pV_\beta(x'+1,1)+(1-p)V_\beta(x'+1,0))\right)\\
&=c-px'+\beta p(V_\beta(1,1)-V_\beta(x'+1,1))\\
&\leq c-px+\beta p(V_\beta(1,1)-V_\beta(x+1,1))\\
&=V_\beta\left((x,1),a=1\right)-V_\beta\left((x,1),a=0\right)\leq 0
\end{aligned}$$

where the above results from the non-decreasing function of $V_\beta(x,\lambda)$ in x for a given λ and β. Similarly we can show that if β-optimal action for state $(x,0)$ is to update, then β-optimal action for state $(x',0)$ is also to update for $x'>x$. Hence, a β-optimal policy is the threshold type and therefore, the cost optimal policy is of the threshold type [35].

Theorem 12: For channels satisfying Assumption 2, if scheduling decisions are made with one-delayed CSI ($D=1$), the sub-problem is Indexable.

Proof:

$$V_\beta(x,1)=\min\{V_\beta((x,1),a=0),V_\beta((x,1),a=1)\};$$

$$x+1+\beta\left(pV_\beta(x+1,1)+(1-p)V_\beta(x+1,0)\right)\ \underset{a=1}{\overset{a=0}{\lessgtr}}\ (1-p)x+1+c+\beta\left(pV_\beta(1,1)+(1-p)V_\beta(x+1,0)\right)$$

$$\Rightarrow px+\beta pV_\beta(x+1,1)\ \underset{a=1}{\overset{a=0}{\lessgtr}}\ c+\beta pV_\beta(1,1)$$

We know that $V_\beta(x,1)$ is non-decreasing in x for a given β. For $c\leq 0$, the optimal action is to update, i.e., $S(0)=\phi$. Note that $\forall c>0,\exists x_0$ such that $px+\beta pV_\beta(x+1,1)>c+\beta pV_\beta(1,1),\forall x>x_0$, i.e., the optimal action is to idle till x_0. As c increases, x_0 increases, so the set $S_2(c)=\{(x,1):x=1,2,...,x_0\}$ monotonically increases. Similarly for states of the form $(x,0)$, as c increases, the set $S_1(c)=\{s(x,0):x=1,2,...\}$ monotonically increases. The union of both these disjoint sets monotonically increases and therefore, the set $S(c)$ monotonically increases to the entire state space as $c\rightarrow\infty$.