

A Whittle Index Approach to Minimizing Functions of Age of Information

Vishrant Tripathi and Eytan Modiano
Laboratory for Information & Decision Systems, MIT

Abstract—We consider a setting where multiple active sources send real-time updates over a single-hop wireless broadcast network to a monitoring station. Our goal is to design a scheduling policy that minimizes the time-average of general non-decreasing cost functions of Age of Information. We use a Whittle index based approach to find low complexity scheduling policies that have good performance. We prove that for a system with two sources, having possibly different cost functions and reliable channels, the Whittle index policy is exactly optimal. We derive structural properties of an optimal policy, that suggest that the performance of the Whittle index policy may be close to optimal in general. These results might also be of independent interest in the study of restless multi-armed bandit problems with similar underlying structure. We further establish that minimizing monitoring error for linear time-invariant systems and symmetric Markov chains is equivalent to minimizing appropriately chosen monotone functions of Age of Information. Finally, we provide simulations comparing the Whittle index policy with optimal scheduling policies found using dynamic programming, which support our results.

I. INTRODUCTION

Many emerging applications depend on the timely delivery of status updates from a number of sources to a central monitor over a single-hop wireless network. Examples include sensor and actuator data for networked control systems, collecting information for IoT applications, mobility data in vehicular networks, and real-time surveillance and monitoring.

Age of Information (AoI) is a metric that captures timeliness of received information at a destination [1], [2]. Unlike packet delay, AoI measures the lag in obtaining information at a destination node, and is therefore suited for applications involving gathering or dissemination of time sensitive updates. Age of information, at a destination, is defined as the time that has elapsed since the last received information update was generated at the source. AoI, upon reception of a new update packet, drops to the time elapsed since generation of the packet, and grows linearly otherwise. Over the past few years, there has been a rapidly growing body of work on analyzing AoI for queuing systems [1]–[6], and using AoI as a metric for scheduling policies in networks [7]–[14].

The problem of minimizing age of information in single-hop networks was first considered in [7] and [8]. In these works, the authors considered a base station collecting time-sensitive information from a number of sources over a wireless broadcast network, where only one source can send an update

at any given time. They looked at weighted linear combinations of AoI of all sources as the metric to be optimized. This prompted the design of low complexity scheduling policies that provably minimize weighted sum AoI at the base station, up to a constant multiplicative factor. These results crucially depend on the fact that for linear AoI, one can find a stationary randomized policy that is factor-2 optimal. As we will see later, this observation does not hold for general functions of AoI. In fact, stationary randomized policies can be arbitrarily worse than simple heuristic policies.

Scheduling problems with weighted linear combinations of age have also been considered with throughput constraints in [9] and with general interference constraints in [10]. AoI-based scheduling with stochastic arrivals was considered in [13], where a Whittle Index policy was shown to have good performance.

On the other hand, nonlinear cost functions of age were introduced as a natural extension to the AoI metric in [2] for characterizing how the level of dissatisfaction depends on data staleness in a more general manner. Nonlinear functions of age of information were also discussed in the context of queuing systems in [15] and [16]. These papers develop the notion of value of information and use nonlinear cost of update delays, which correspond to nonlinear age cost functions.

Nonlinear functions of age have also been discussed in the context of networked control systems in [17], [18] and [19]. In [17], the authors discuss a real time networked control system and show that the cost function is characterized as a non-decreasing, possibly nonlinear, function of AoI. In [18], the authors formulated the state estimation problem for an LTI system, where the state of a discrete-time LTI system can be observed in any time-slot by paying a fixed transmission cost. The problem of minimizing the time-average of the sum of the estimation error and transmission cost reduces to minimizing a non-decreasing age-cost function for a single source with a fixed transmission cost. We explore this relationship more closely in Section VII, where we establish a similar equivalence for monitoring multiple LTI systems.

In this work, we consider a setting similar to the one in [7] and [8]. We look at wireless broadcast network with N sources generating real-time updates that need to be sent to a monitoring station. In any time-slot, only one source can attempt a transmission to the base station. Instead of weighted sum AoI, we are interested in minimizing the time-average of *general non-decreasing cost functions* of AoI, summed over all sources. Examples of such functions include $f(x) = 2^x$,

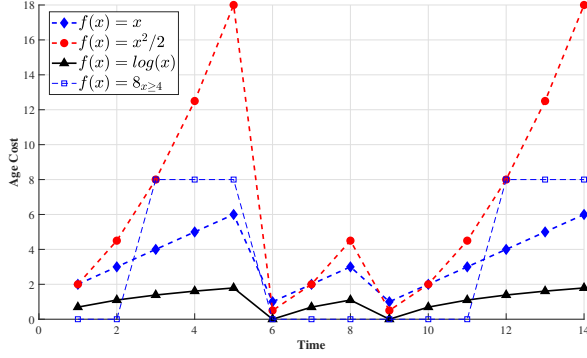


Fig. 1: Linear, quadratic, logarithmic and indicator cost functions for a sample age process. The linear process tracks the actual values of AoI.

$f(x) = \log(x)$, $f(x) = \mathbb{1}_{\{x \geq 10\}}$, etc. See Fig.1 for examples. We develop a restless multi-armed bandit formulation for the problem and use a Whittle Index based approach to find low complexity scheduling policies that have good performance.

Scheduling to minimize functions of age has also been considered in [11] and [12]. In [11], the authors deal with minimizing symmetric functions of age of sources over multiple orthogonal unreliable channels and show that simple greedy schemes are asymptotically optimal. In [12], the authors formulate the general functions of age problem with reliable channels and develop a high complexity algorithm that achieves minimum age. They also derive a key structural property of the optimal policy in this setting - the optimal policy is always periodic. However, their approach does not extend to the setting with unreliable channels. In this work, we consider unreliable channels and also build upon results from [12] and [13] to derive stronger structural properties for optimal policies. These properties hint at why the performance of the heuristic Whittle index policy may be close to optimal. Moreover, it has been shown recently that the Whittle policy is indeed asymptotically optimal for *linear* functions of AoI [20].

The remainder of the paper is organized as follows. In Section II, we describe the general system model. In Section III, we describe the equivalent restless multi-armed bandit formulation and discuss why we use the Whittle Index approach to solve the problem. In Section IV, we discuss the functions of age problem with reliable channels, develop the Whittle Index solution for this setting, and also prove key structural properties that an optimal policy must satisfy. In Section V, we find the Whittle Index policy for the functions of age problem with unreliable channels. In Section VI, we provide simulation results that verify our theoretical results. In Section VII, we show that the problem of minimizing monitoring error for linear time-invariant systems when observing them over a wireless channel is equivalent to minimizing functions of AoI. We also show a similar result for monitoring symmetric Markov chains over a wireless channel. This shows the direct applicability of our Whittle framework to a large class of

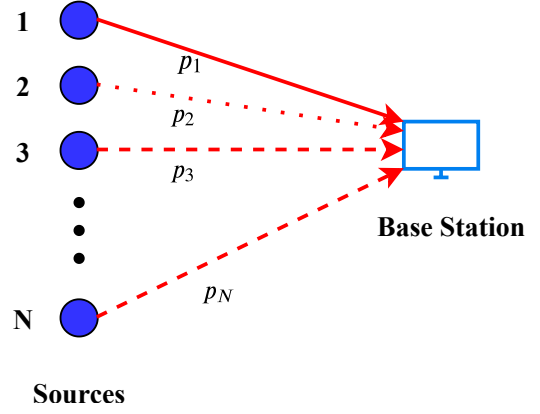


Fig. 2: N sources transmitting updates to a base station over a wireless channel, with different reliabilities.

wireless monitoring problems.

A preliminary version of this paper appeared in the conference proceedings of Allerton 2019 [21].

II. MODEL

Consider a single-hop wireless network with N active sources generating real-time status updates that need to be sent to a base station. We consider a slotted system in which each source takes a single time-slot to transmit an update to the base station. Due to interference, only one of the sources can transmit in any given time-slot.

For every source i , the age of information at the base station $A_i(t)$ measures the time elapsed since it received a fresh information update from the source. We assume active sources, i.e. in any time-slot, sources can generate fresh updates at will. Let $s(t)$ be the source activated in time-slot t and $u_i(t)$ be a Bernoulli random variable with parameter p_i that denotes channel reliability between the i^{th} source and the base station. Then, we have

$$A_i(t+1) = \begin{cases} A_i(t) + 1, & \text{if } s(t) \neq i \text{ or } u_i(t) = 0, \\ 1, & \text{if } s(t) = i \text{ and } u_i(t) = 1. \end{cases} \quad (1)$$

In this work, we consider general cost functions of age as our metric of interest. For each source i , let $f_i(\cdot)$ denote a positive *non-decreasing* cost function.

Let π be a scheduling scheme that decides which sources to schedule in every time-slot. The age process $A_i(t)$ depends on π and the channel processes. Then, the expected average cost of age for source i is given by

$$C_i^{\text{ave}}(\pi) \triangleq \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{t=1}^T f_i(A_i^\pi(t)) \right], \quad (2)$$

where $A_i^\pi(t)$ is age process for the i^{th} source under policy π .

Our goal is find a schedule π that minimizes the sum of average costs of age of sources, i.e. (2). Let Π denote the

set of causal scheduling policies, then we want to solve the following optimization problem

$$C^* = \min_{\pi \in \Pi} \sum_{i=1}^N C_i^{\text{ave}}(\pi), \quad (3)$$

where C^* is minimum average cost and π^* is an optimal scheduling policy.

III. RESTLESS MULTI-ARMED BANDIT FORMULATION

The restless multi-armed bandit (RMAB) is a classical resource allocation problem that has been studied in the optimization and operations research community for many decades. It involves N Markov bandits, each of which evolves according to two different transition laws - one for when it is active, and one for when it is not. The scheduler can only activate one arm at any given time-slot, and a cost function maps the states of the arm to a corresponding cost in each time-slot. The goal of the scheduler is to design an arm activation policy that minimizes the long term time-average cost. The general solution strategy for such resource allocation problems is dynamic programming. However, that suffers from the curse of dimensionality and is not computationally feasible. Whittle, in his seminal work [22], showed that RMABs admit low complexity heuristic solutions called the Whittle Index if they satisfy a special property called indexability. In this section, we will show that scheduling to minimize such a metric can be reformulated as a RMAB.

Consider a restless multi-armed bandit problem with N arms. The state space for every arm i is the set of positive integers \mathbb{Z}^+ . The state evolution of the arm depends on whether it is currently active or not. Let the state of arm i at time t be denoted by $A_i(t)$. If arm i is active in time-slot t then the state evolution is given by

$$A_i(t+1) = \begin{cases} A_i(t) + 1, & \text{w.p. } 1 - p_i \\ 1, & \text{w.p. } p_i. \end{cases} \quad (4)$$

If the arm is not active in time-slot t , then the state evolution is given by

$$A_i(t+1) = A_i(t) + 1. \quad (5)$$

For every arm i , there is a cost function $f_i : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ which maps the states of the arm to their associated costs. Thus, the cost of a state $\mathbf{x} \in \mathbb{Z}^{+N}$ is given by $\sum_{i=1}^N f_i(x_i)$, where \mathbf{x} is a vector of states (ages) and x_i is the state (age) of the i^{th} source. Given that only one arm can be activated in any time-slot, the goal of the RMAB framework is to find a scheduling policy that minimizes the total time average cost of running this system.

This establishes the equivalence between the functions of age problem discussed earlier and a corresponding restless multi-armed bandit. Observe that the “restless” part of our construction cannot be dropped, since the states of the arms do not freeze when they are not active and there is no way to reformulate our problem as a simple (non-restless) multi-armed bandit problem. If that were the case, we could have

found an optimal policy by solving for the Gittins index [23]. However, finding optimal policies for restless bandits is much harder. The usual approach is to find the Whittle Index policy which provides good performance under certain conditions, namely *indexability* of the RMAB problem.

In [7] and [8], the authors develop three methods to solve the minimum age scheduling problem. First, they look at stationary randomized policies, where a source i is scheduled at random with a fixed probability p_i . They find a stationary randomized policy that is factor-2 optimal for weighted sum AoI. However, this result does not hold for general functions: even the best stationary randomized policies in our setting can lead to an unbounded overall cost, despite there being very simple policies that have bounded cost. We demonstrate this with a simple example.

Consider two identical sources with cost functions given by $f(x) = 3^x$ and reliable channels, i.e $p_1 = p_2 = 1$. Any stationary randomized policy schedules at least one of the sources with probability less than or equal to 0.5. For this source, the average cost is lower bounded by $\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T (3^t)^{\frac{0.5^T}{T}}$ since with probability at least 0.5, it does not get to transmit and its age increases by 1 in every time-slot. Observe that this lower bound goes to ∞ and hence the average cost also goes to ∞ for all stationary randomized policies. On the other hand, a simple round-robin scheme that schedules the two sources in alternating time-slots guarantees bounded cost for both sensors. Thus, stationary randomized policies can be infinitely worse than the optimal policy for the functions of age problem.

The second method developed for age-based scheduling in [7], [8] uses a Max-Weight approach. The authors design a quadratic Lyapunov function for the weighted sum of linear functions of AoI and find the max-weight policy - the policy that maximizes the amount of negative drift in the Lyapunov function in every time-slot. Performance guarantees for the max-weight policy crucially rely on the fact that there exists a stationary randomized policy that is factor-2 optimal for linear functions of age. Since this is not the case for general functions of age, we cannot develop similar performance bounds using a Max-Weight policy for the general functions of age problem.

This finally leaves us with the third method - using a Whittle Index based approach. In the following two sections, we use the RMAB formulation to establish indexability for the functions of age problem and derive a Whittle Index policy. We also show that for the case with 2 sources and reliable channels, the Whittle index policy is exactly optimal. This is a novel result since the optimality of Whittle Index policies is typically shown either only asymptotically, or in symmetric settings for finite systems. On the other hand, our optimality result holds for two asymmetric sources.

IV. RELIABLE CHANNELS

We first look at the problem with reliable channels between the sources and the base station. This leads to simpler analysis and a better understanding of the problem. Consider the setup described in Section I with channel reliability $u_i(t) = 1$, for all i and t . In other words, the probability of success $p_i = 1, \forall i$.

In Section III, we showed that the functions of age minimization problem is equivalent to a restless multi-armed bandit problem. Next, we use a Whittle Index based approach to try and solve the problem.

The first step in the Whittle Index approach is to formulate the *decoupled problem*, where we consider a single arm in isolation with a fixed charge required to activate the arm.

Definition Decoupled Problem

Consider a single arm with the state space \mathbb{Z}^+ and an associated non-decreasing cost function $f : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$. Let the state of the arm be $A(t)$. Its evolution is given by

$$A(t+1) = \begin{cases} A(t) + 1, & \text{if not active at time } t \\ 1, & \text{otherwise.} \end{cases}$$

There is a strictly positive activation charge C to be paid in every time-slot that the arm is pulled.

Our goal is to find a scheduling policy that minimizes the time-average cost of running this system. Assuming that the cost function $f(\cdot)$ is non-negative and non-decreasing, we solve the decoupled problem using dynamic programming. The case when the activation charge is set to zero is trivial. The optimal policy is to always activate the arm. So, we consider C to be strictly positive. The single source decoupled problem has also been solved in a slightly different setting in [18].

Theorem 1: The optimal policy for the decoupled problem is a stationary threshold policy. Let H satisfy

$$f(H) \leq \frac{\sum_{j=1}^H f(j) + C}{H} \leq f(H+1). \quad (6)$$

Then, the optimal policy is to activate the arm at time-slot t if $A(t) \geq H$ and to let it rest otherwise. If no such H exists, the optimal policy is to never activate the arm.

Proof: See Appendix A. ■

Theorem 1 establishes that the optimal policy for the decoupled problem has a threshold structure. We now want to show that the *indexability* property also holds for the decoupled problem. The indexability property states that as the activation charge C increases from 0 to ∞ , the set of states for which it is optimal to activate the arm decreases monotonically from the entire set \mathbb{Z}^+ to the empty set $\{\phi\}$.

Theorem 2: The *indexability* property holds for the decoupled problem.

Proof: See Appendix B. ■

The Whittle index approach states that if the decoupled problem satisfies the indexability property, we can formulate a heuristic index policy called the Whittle Index Policy that

has good performance.

Definition Whittle Index

Consider the decoupled problem and denote by $W(h)$ the Whittle index in state h . Given indexability, $W(h)$ is the infimum charge C that makes both decisions (activate, not activate) equally desirable in state h . The expression for $W(h)$ is given by

$$W(h) = hf(h+1) - \sum_{j=1}^h f(j). \quad (7)$$

Observe that using (6), $C = W(h)$ is the minimum value of the activation charge that makes both actions equally desirable in state h . This gives us the expression for the Whittle index.

Let $W_i(x) := xf_i(x+1) - \sum_{j=1}^x f_i(j)$ represent the index function for the i^{th} decoupled problem. By the monotonicity of $f_i(\cdot)$, it is easy to see that the functions $W_i(\cdot)$ are also monotonically non-decreasing. This is because $W_i(h) - W_i(h-1) = h(f_i(h+1) - f_i(h)) \geq 0, \forall h$ since $f_i(\cdot)$ is non-decreasing. Using these functions, we define the Whittle Index Policy.

Definition Whittle Index Policy

Let $\pi^W(t)$ be the action taken by the Whittle Index Policy at time t . Then $\pi^W(t)$ is given by

$$\begin{aligned} \pi^W(t) &= \arg \max_{1 \leq i \leq N} \left\{ W_i(A_i(t)) \right\} \\ &= \arg \max_{1 \leq i \leq N} \left\{ A_i(t)f_i(A_i(t)+1) - \sum_{j=1}^{A_i(t)} f_i(j) \right\}. \end{aligned} \quad (8)$$

Consider the case when the cost functions are weighted linear functions of AoI, i.e let $f_i(A_i(t)) = w_i A_i(t)$, with positive weights w_i . This is the setting considered in [7] and [8]. The Whittle Index for source i is then given by $W_i(A_i(t)) = w_i(A_i^2(t) + A_i(t))/2$. This is the same as the Whittle index found in [7], where the authors showed that the Whittle policy is optimal for symmetric settings when all the weights are equal. We also establish that for $N = 2$, the Whittle index policy is optimal even for asymmetric settings.

Theorem 3: For the functions of age problem with reliable channels and two sources, the Whittle index policy is exactly optimal.

Proof: See Appendix F. ■

This is an atypical result for restless multi-armed bandit problems which typically only have optimality results for symmetric or asymptotic settings. Our result is valid for finite

($N = 2$) asymmetric settings. To the best of our knowledge, this is the first work to prove such a result for a restless multi-armed bandit problem. Next, we discuss some general properties that an optimal policy satisfies even for larger size systems. These properties help us establish the optimality of the Whittle index policy for $N = 2$ and provide insight as to why the Whittle index policy has good performance in general.

A. Properties of an Optimal Policy

For the functions of age problem, a policy is stationary if it depends only on the current values of age. A cyclic policy is one that repeats a finite sequence of actions in a fixed order. We define the space of policies that are stationary and periodic.

Definition Stationary Cyclic Policies

A stationary cyclic policy is a stationary policy that cycles through a finite subset of points in the state space, repeating a fixed sequence of actions in a particular order.

In [12], the authors show that for reliable channels there exists an optimal policy that is stationary, cyclic and can be found by solving the minimum average cost cycle problem over a large graph.

We look at this cyclic policy and analyze its properties. If there are multiple such cycles, we consider a cycle with the shortest length. We denote the length of the cycle by T and age vectors on the cycle to be $\mathbf{x}_1, \dots, \mathbf{x}_T$. Let the corresponding scheduling decisions be d_1, \dots, d_T . This implies that for state \mathbf{x}_k , taking action d_k leads to the state \mathbf{x}_{k+1} , where the subscripts cycle back to $1, 2, \dots$ after T .

We establish an important structural property that such an optimal policy must satisfy, which we call the *strong-switch-type* property. We call the policies that satisfy this property *strong-switch-type* policies.

Definition Strong-switch-type Policies

Consider a stationary policy π that maps every point in the state space \mathbb{Z}^{+N} to the set of arms $\{1, \dots, N\}$. We say that such a policy is strong-switch-type if

$$\pi(x_1, \dots, x_N) = i$$

implies

$$\pi(x'_1, \dots, x'_N) = i,$$

for all \mathbf{x} and \mathbf{x}' such that $x'_i \geq x_i$ and $x'_j \leq x_j, \forall j \neq i$.

In words, the strong-switch-type property implies that if a policy decides to activate arm i for a state vector \mathbf{x} , then for a state vector \mathbf{x}' with a higher age for the i^{th} source and lower ages for all the other sources, it still decides to activate source i . Note that our definition of strong-switch-type policies is a stronger version of the switch-type policies introduced in [13].

Theorem 4: For the functions of age problem with reliable channels, all state-action pairs that are a part of the shortest length optimal cyclic policy must satisfy the strong-switch-type property.

Proof: See Appendix C. ■

We can prove this result for general values of N . However, to extend the strong-switch-type property over the entire state-space, we consider systems with up to three sources.

Theorem 5: There exists an optimal stationary policy for the functions of age problem with reliable channels and up to three sources that has the strong-switch-type property over the entire state-space.

Proof: We have already established that points on the minimum average cost cycle satisfy the strong-switch-type property. In Appendix D, we extend this policy over the entire state space while maintaining the strong-switch property to obtain a well defined stationary policy. ■

While we prove this result for up to three source and reliable channels, we believe that the strong-switch-type property is a natural property that an optimal policy must have in general, due to monotonicity of cost functions.

We now define the space of policies that can be found as a result of the Whittle Index based approach.

Definition Index Policies

Consider a stationary policy π that maps every point in the state space \mathbb{Z}^{+N} to the set of arms $\{1, \dots, N\}$. We say that such a policy is an index policy if

$$\pi(x_1, \dots, x_N) = \arg \max_{1 \leq i \leq N} \left\{ F_i(x_i) \right\}$$

for all \mathbf{x} , where $F_i : \mathbb{Z}^+ \rightarrow \mathbb{R}$ are monotonically non-decreasing functions for all i .

Observe that if F_i are the same as W_i in the above definition, then we get back the Whittle Index Policy. Also, note that an index policy always satisfies the strong-switch-type property by definition. This is because the index functions $F_i(\cdot)$ are monotonically non-decreasing. We now show that index policies are in fact the same as strong-switch-type policies.

Theorem 6: For the functions of age problem, every policy that is strong-switch-type is also an index policy.

Proof: The proof is based on induction on the number of sources. We assume that every strong-switch-type policy can be represented as an index policy for systems with N

sources. Using this fact, we show that strong-switch-type policies can also be represented as index policies for systems with $N+1$ sources. We also show that the two types of policies are equivalent for the single source decoupled problem, thus completing the proof. The details are in Appendix E. ■

An important point to notice is that while we use the reliability of channels in the proof of Theorem 5, we do not use any such condition for the proof of Theorem 6. Thus, strong-switch-type policies are equivalent to index policies regardless of channel connectivity.

Theorems 5 and 6 together imply the following corollary.

Corollary 1: For the functions of age problem with reliable channels and up to three source, there exists a stationary optimal policy that is an index policy.

In other words, there exists an optimal policy that looks like the Whittle Index policy in that the arm to be activated has the maximum value among *monotone index functions* that take as arguments only the states of individual arms. This hints at why the performance of Whittle Index policies may be close to optimal.

Observe that the Whittle Index policy would be optimal in general if we could show that it achieves a cost that is the minimum cost among the space of index policies and that the strong-switch-type property holds for some optimal policy. We show that this is indeed the case for $N = 2$. However, we later provide an example that shows that the Whittle policy is not optimal, but only close to optimal, for $N = 4$.

We leave the question of whether the Whittle index policy is at most a constant factor away from optimal in general to future work. We believe that the structural properties introduced here provide a recipe to proving constant factor optimality of the Whittle index policy, even for general bandit problems with similar underlying structure.

V. UNRELIABLE CHANNELS

We now consider independent Bernoulli channels between every source and the base station, with probability of success p_i for source i . We derive a Whittle index in this setting and establish indexability of the RMAB problem by enforcing a bounded cost condition on the functions $f_i(\cdot)$.

An important fact to notice is that monotonicity in itself is not sufficient to ensure that the system has finite average cost even for $N = 1$, in the case of unreliable channels. Consider a single source case where $f(a) = 3^a$ and the probability of success $p = 0.5$. If the source attempts a transmission in every time-slot, the expected average cost satisfies

$$\limsup_{T \rightarrow \infty} \sum_{t=1}^T (3^t) \frac{0.5^T}{T} \leq \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{t=1}^T 3^{A(t)} \right], \quad (9)$$

since with probability 0.5, the transmission fails and age increases by 1 in every time-slot. However, observe that the summation on the left goes to infinity and thus the expected average cost goes to infinity. This happens despite the source

attempting a transmission in every time-slot. To prevent such a situation from happening we enforce the following *bounded cost* condition on the age cost functions f_i in addition to monotonicity

$$\sum_{h=1}^{\infty} f_i(h)(1-p_i)^h < \infty. \quad (10)$$

It can be shown that this condition ensures that the single source case has bounded cost. We define the decoupled problem in this case as follows:

Definition Decoupled Problem

Consider a single arm with the state space \mathbb{Z}^+ , probability of success p and an associated non-decreasing cost function $f : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ that satisfies the *bounded cost* condition. Let the state of the arm be $A(t)$. If the arm is active at time t , its evolution is given by

$$A(t+1) = \begin{cases} A(t) + 1, & \text{w.p. } 1-p \\ 1, & \text{w.p. } p. \end{cases}$$

If the arm is not active in time-slot t , then the state evolution is given by

$$A(t+1) = A(t) + 1.$$

There is a strictly positive activation charge C to be paid in every time-slot that the arm is pulled.

As before, our goal is to find a scheduling policy that minimizes the time-average cost of running this system.

Theorem 7: The optimal policy for the decoupled problem is a stationary threshold policy. Let H satisfy

$$\begin{aligned} p^2(H-1) \left(\sum_{k=H}^{\infty} f(k)(1-p)^{k-H} \right) - p \left(\sum_{j=1}^{H-1} f(j) \right) \\ \leq C \\ \leq p^2 H \left(\sum_{k=H+1}^{\infty} f(k)(1-p)^{k-H-1} \right) - p \left(\sum_{j=1}^H f(j) \right) \end{aligned} \quad (11)$$

Then, the optimal policy is to activate the arm at time-slot t if $A(t) \geq H$ and to let it rest otherwise. If no such H exists, the optimal policy is to never activate the arm.

Proof: See Appendix G. ■

Observe that taking the limit as $p \rightarrow 1$ in Theorem 7, we get back the threshold policy for reliable channels derived in Theorem 1. We now establish indexability and derive the functional form of the Whittle Index.

Theorem 8: The *indexability* property holds for the decoupled problem. Denote by $W(h)$ the Whittle index in state h . Given indexability, $W(h)$ is the infimum charge C that makes both decisions (activate, not activate) equally desirable in state h . The expression for $W(h)$ is given by

$$W(h) = p^2 h \left(\sum_{k=1}^{\infty} f(k+h)(1-p)^{k-1} \right) - p \left(\sum_{j=1}^h f(j) \right). \quad (12)$$

Proof: See Appendix H. ■

Again, observe that taking the limit as $p \rightarrow 1$, we get back the Whittle Index derived in Section IV. Further, if we assume that the cost functions are weighted linear functions of AoI, i.e. $f_i(A_i(t)) = w_i A_i(t)$ where all the weights are positive, then the index functions for the Whittle policy are given by $W_i(A_i(t)) = w_i p_i A_i(t) (A_i(t) + \frac{1+(1-p_i)}{1-(1-p_i)})/2$. This corresponds to the index policy developed in [7], where the authors showed that for symmetric settings when all the weights and channels probabilities are equal, the Whittle index policy is optimal.

VI. SIMULATIONS

First, we compare the optimal policy, found using dynamic programming, with the Whittle index policy for two sources. We consider six different settings in total - 3 sets of functions, each with reliable and unreliable channels.

For settings A_1 and A_2 , the cost functions are chosen to be $f_1(x) = 13x$ and $f_2(x) = x^2$. In A_1 , we consider reliable channels, i.e. $p_1 = p_2 = 1$. In A_2 , we consider unreliable channels, specifically $p_1 = 0.9$ and $p_2 = 0.5$. For settings B_1 and B_2 , the cost functions are chosen to be $f_1(x) = x^2$ and $f_2(x) = 3x$. In B_1 , we consider reliable channels, i.e. $p_1 = p_2 = 1$. In B_2 , we consider unreliable channels, specifically $p_1 = 0.65$ and $p_2 = 0.8$. For settings C_1 and C_2 , the cost functions are chosen to be $f_1(x) = x^3/2$ and $f_2(x) = 10 \log(x)$. In C_1 , we consider reliable channels, i.e. $p_1 = p_2 = 1$. In C_2 , we consider unreliable channels, specifically $p_1 = 0.55$ and $p_2 = 0.75$. Simulation results are presented in Table I.

Setting	Optimal Cost	Whittle Index Cost
A_1 (reliable)	21.95	21.95
A_2 (unreliable)	36.12	36.28
B_1 (reliable)	8.48	8.48
B_2 (unreliable)	23.16	23.37
C_1 (reliable)	5.69	5.69
C_2 (unreliable)	21.54	21.54

TABLE I: Cost of the Whittle index policy and the optimal dynamic programming policy for 2 sources.

We find the optimal cost for each setting using finite horizon dynamic programming over a horizon of 500 time-slots. For reliable channels, we find the cost of the Whittle index policy by simply implementing it once over 500 time-slots. For

unreliable channels, we estimate the expected Whittle index cost by averaging the performance of the Whittle index policy over 500 independent runs.

Observe that the Whittle index policy is exactly optimal when the channels are reliable, as expected from our theoretical results. The expected cost for the Whittle index policy is very close to the optimal cost for unreliable channels as well. Also, for the same set of functions, having unreliable channels increases the cost compared to reliable channels, as expected.

Next, we compare the optimal policy with the Whittle index policy for more than two sources. Simulation results are presented in Table II.

For settings D_1 and D_2 , we consider 3 sources. The cost functions are chosen to be $f_1(x) = x^2$, $f_2(x) = 3x$ and $f_3(x) = x^4$. In D_1 , we consider reliable channels, i.e. $p_1 = p_2 = p_3 = 1$. In D_2 , we consider unreliable channels, specifically $p_1 = 0.66$, $p_2 = 0.8$ and $p_3 = 0.75$.

For settings E_1 and E_2 , we consider 4 sources. The cost functions are chosen to be $f_1(x) = x^3$, $f_2(x) = 2x$, $f_3(x) = 15x$ and $f_4(x) = x^2$. In E_1 , we consider reliable channels, i.e. $p_1 = p_2 = p_3 = 1$. In E_2 , we consider unreliable channels, specifically $p_1 = 0.7$, $p_2 = 0.9$, $p_3 = 0.67$ and $p_4 = 0.8$.

No. of Sources	Setting	Optimal Cost	Whittle Index Cost
3	D_1 (reliable)	44.23	44.23
	D_2 (unreliable)	161.19	161.39
4	E_1 (reliable)	73.36	73.36
	E_2 (unreliable)	129.02	130.94
4	F_1 (reliable)	87.66	88.27
	F_2 (unreliable)	158.35	159.81

TABLE II: Cost of the Whittle index policy and the optimal policy for more than 2 sources.

For settings F_1 and F_2 , we consider 4 sources. The cost functions are chosen to be $f_1(x) = x^3$, $f_2(x) = e^x$, $f_3(x) = 15x$ and $f_4(x) = x^2$. In F_1 , we consider reliable channels, i.e. $p_1 = p_2 = p_3 = 1$. In F_2 , we consider unreliable channels, specifically $p_1 = 0.8$, $p_2 = 0.85$, $p_3 = 0.75$ and $p_4 = 0.66$.

We observe that the cost of the Whittle index policy is the same as that obtained using dynamic programming for settings D_1 and E_1 . However, for setting F_1 , we observe a small gap in performance between the two policies, thus giving us an example that shows that the *Whittle index policy need not be optimal*, in general. We also verify that the optimal policy found using dynamic programming follows a cyclic pattern that satisfies the strong-switch-type property and is distinct from the Whittle index policy. This is also in line with our discussion on structural properties.

We note that computing the optimal policy using dynamic programming becomes progressively harder in terms of space and time complexity for larger values of N , as the state-space to be considered grows exponentially with N . The Whittle index policy, on the other hand, is very easy to compute and implement with only a linear increase in space and time complexity with the number of sources. Also, as is evident from simulations, the performance of the Whittle policy is

close to optimal in every setting considered, thus making it a very good low complexity heuristic.

VII. APPLICATIONS

In this section, we will apply the framework we have developed to two problems in remote monitoring and control to show that optimizing general functions of AoI arise naturally in many practical settings.

A. Monitoring LTI systems

First, we consider the remote monitoring of linear time-invariant (LTI) systems over a wireless channel. Suppose that there are N such systems, where the i th system evolves over time as follows

$$x_i(t+1) = G_i x_i(t) + w_i(t), \quad (13)$$

where $x_i(t) \in \mathbb{R}^{d_i}$, $G_i \in \mathbb{R}^{d_i \times d_i}$ is the system matrix and $w_i(t) \sim \mathcal{N}(0, \Sigma_i)$ is multi-variate zero-mean Gaussian noise, i.i.d. across time. We further assume that the noise increments $w_i(t)$ are independent across sources, so their evolution is decoupled.

Suppose that a central agent wants to monitor the state of each of the N systems with as little monitoring error as possible. However, due to wireless interference constraints, it can only observe the state of one system at any given time-slot. How should the agent design a wireless scheduling policy that minimizes expected monitoring error?

Let $\hat{x}_i(t)$ represent the maximum likelihood estimate of the state of the i th system at the monitor at any given time-slot t , given past observations. We define monitoring error for the i th system as

$$e_i(t) \triangleq \mathbb{E} \left[\|x_i(t) - \hat{x}_i(t)\|_2^2 \right]. \quad (14)$$

The following theorem relates the expected monitoring error of the i th system to its AoI. Specifically, we compute the expected error if the i th system has not been observed for the last Δ time-slots.

Theorem 9: Suppose that the i th system evolves according to (13). Further suppose that the monitor last observed the state of the i th system at time $t = \tau$. Then, the expected monitoring error for the i th system at time $t = \tau + \Delta$ is given by

$$e_i(\tau + \Delta) = \mathbb{E} \left[\|x_i(\tau + \Delta) - \hat{x}_i(\tau + \Delta)\|_2^2 \right] = \sum_{k=0}^{\Delta-1} \text{Tr}((G_i^k)^T (G_i^k) \Sigma_i) \triangleq f_i(\Delta). \quad (15)$$

Proof: See Appendix I. ■

Using this observation, we can establish an equivalence between minimizing monitoring error and minimizing functions of AoI. To find the scheduling policy π that minimizes

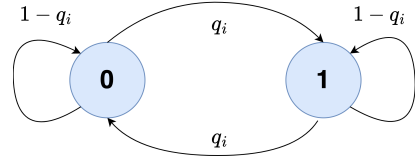


Fig. 3: Symmetric two-state Markov chain, representing the state of the i th process.

expected time-average monitoring error, we need to solve the following optimization problem

$$\min_{\pi \in \Pi} \limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{t=1}^T \sum_{i=1}^N e_i(t) \right], \quad (16)$$

where $e_i(t)$ is defined as in (14). This optimization problem is equivalent to solving the following functions of AoI problem

$$\min_{\pi \in \Pi} \limsup_{T \rightarrow \infty} \frac{1}{T} \left[\sum_{t=1}^T \sum_{i=1}^N f_i(A_i(t)) \right], \quad (17)$$

where $A_i(t)$ is the AoI of the i th system and the functions $f_i(\cdot)$ are as defined in (15).

We also show in Appendix I that the functions $f_i(\cdot)$ are monotonically increasing, so we can indeed apply our Whittle index approach to solve this problem. The rate at which the functions $f_i(\cdot)$ increase depends on the eigenvalues of the system matrices G_i . If the largest eigenvalue of G_i lies inside (outside) the unit circle, then $f_i(\cdot)$ increases slower (faster) than a linear function. If the largest eigenvalue of G_i lies on the unit circle, then $f_i(\cdot)$ increases linearly.

B. Monitoring Markov Chains

Consider N symmetric two state Markov chains of the form drawn in Fig. 3 running in discrete-time. As for the previous example, we assume that only one system out of the N can be observed in any given time-slot. We denote the distribution of the i th Markov chain at time t by $x_i(t)$, where $x_i(t) = [1 \ 0]$ if the Markov chain is in state 0 and $x_i(t) = [0 \ 1]$ if the Markov chain is in state 1.

We assume that the base station knows the transition probability q_i and the transition matrix

$$Q_i = \begin{bmatrix} 1 - q_i & q_i \\ q_i & 1 - q_i \end{bmatrix}$$

associated with the i th Markov chain and uses this to maintain the estimated distribution of the i th chain, based on the most recent observation. Suppose that the base station knew that the i th Markov chain had the distribution $x_i(\tau)$ at time τ . Using the transition matrix Q_i for the i th chain, the base station can compute the distribution of the Markov chain at time $\tau + \Delta$ given the information at time τ . We denote this estimated distribution of the actual state by $\hat{x}_i(\tau + \Delta)$ and it is given by

$$\hat{x}_i(\tau + \Delta) = x_i(\tau) \begin{bmatrix} 1 - q_i & q_i \\ q_i & 1 - q_i \end{bmatrix}^\Delta = x_i(\tau) Q_i^\Delta. \quad (18)$$

We are interested in minimizing the monitoring error, defined as a notion of distance between the estimated distribution

and the actual state of the Markov chain. We define error for the i th system as follows -

$$e_i(t) = \mathbb{E} \left[D(x_i(t) || \hat{x}_i(t)) \right], \quad (19)$$

where D is a notion of divergence between the two probability distributions. In this work, we will discuss our results for Kullback-Liebler (KL) divergence and total variation (TV) distance, however, the general ideas should work for other divergences as well. The KL divergence for discrete distributions is defined as

$$D_{KL}(P||Q) = - \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right).$$

The total variation (TV) distance for discrete distributions is defined as

$$D_{TV}(P||Q) = \sum_{x \in \mathcal{X}} \frac{1}{2} |P(x) - Q(x)|.$$

The following theorem relates the expected monitoring error of the i th system to its AoI. Specifically, we compute the expected error if the i th system has not been observed for the last Δ time-slots.

Theorem 10: Suppose that the i th system evolves according to Markov chain in Fig. 3. Further suppose that the monitor last observed the state of the system at time $t = \tau$. Then, the expected monitoring error for the i th system at time $t = \tau + \Delta$ is given by

$$\begin{aligned} e_i(\tau + \Delta) &= \mathbb{E} \left[D(x_i(\tau + \Delta) || \hat{x}_i(\tau + \Delta)) \right] \\ &= \begin{cases} H([Q_i^\Delta]_{00}), & \text{if } D \text{ is KL Divergence,} \\ 2[Q_i^\Delta]_{00}(1 - [Q_i^\Delta]_{00}), & \text{if } D \text{ is TV Distance.} \end{cases} \\ &= f_i(\Delta). \end{aligned} \quad (20)$$

Here $[Q_i^\Delta]_{00}$ is the top diagonal element of the transition matrix raised to the power Δ , i.e. Q_i^Δ and $H(q) \triangleq -q \log(q) - (1 - q) \log(1 - q)$ is the binary entropy function.

Proof: See Appendix J. ■

Using the result above, it is straightforward to establish an equivalence between minimizing monitoring error for Markov chains and minimizing functions of AoI. As for the case with the LTI systems, we further show in Appendix J that the functions $f_i(\cdot)$ are monotonically increasing, so we can indeed apply our Whittle index approach to solve this problem.

An interesting observation for the KL divergence case is that the monitoring error cost ends up being the entropy of the estimated distribution of the Markov chain. This can be interpreted as the amount of uncertainty that the base station has about the Markov chain, which increases with the number of time-slots that the chain remains unobserved. We use this Markov model and our Whittle framework to solve a robotics

problem involving time-varying multi-agent occupancy grid mapping in [24].

VIII. CONCLUSION

In this work, we presented the problem of minimizing functions of age of information over a wireless broadcast network. We used a restless multi-armed bandit approach to establish indexability of the problem and found the Whittle index policy. For the case with two sources and reliable channels, we were able to show that the Whittle index policy is optimal. We also established structural properties of an optimal policy, for the case with reliable channels. These properties hint at why the performance of the Whittle index policy is close to optimal in general.

A possible direction of future work is to try and prove constant factor optimality of the Whittle index policy in general, using the structural properties developed in this work. Other interesting extensions could be to consider sources with stochastic arrivals instead of active sources and handling unknown, possibly time-varying functions of Age of Information.

REFERENCES

- [1] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?," in *Proc. INFOCOM*, pp. 2731–2735, 2012.
- [2] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, pp. 7492–7508, Nov. 2017.
- [3] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Minimizing the age of information through queues," *IEEE Transactions on Information Theory*, 2019.
- [4] L. Huang and E. Modiano, "Optimizing age-of-information in a multi-class queueing system," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 1681–1685, IEEE, 2015.
- [5] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," *arXiv preprint arXiv:1804.06139*, 2018.
- [6] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "On the age of information with packet deadlines," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6419–6428, 2018.
- [7] I. Kadota, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Minimizing the age of information in broadcast wireless networks," in *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 844–851, IEEE, 2016.
- [8] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 26, no. 6, pp. 2637–2650, 2018.
- [9] I. Kadota, A. Sinha, and E. Modiano, "Scheduling algorithms for optimizing age of information in wireless networks with throughput constraints," *IEEE/ACM Transactions on Networking*, 2018.
- [10] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," in *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 61–70, ACM, 2018.
- [11] V. Tripathi and S. Moharir, "Age of information in multi-source systems," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2017.
- [12] P. R. Jhunjunwala and S. Moharir, "Age-of-information aware scheduling," in *SPCOM*, 2018.
- [13] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals: The no-buffer case," *arXiv preprint arXiv:1712.07419*, 2017.
- [14] S. Farazi, A. G. Klein, J. A. McNeill, and D. R. Brown, "On the age of information in multi-source multi-hop wireless status update networks," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, IEEE, 2018.

- [15] 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)*, pp. 326–330, IEEE, 2017.
- [16] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing," *arXiv preprint arXiv:1812.09320*, 2018.
- [17] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using aoi in a single-loop networked control system," *arXiv preprint arXiv:1901.06694*, 2019.
- [18] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "Aoi-penalty minimization for networked control systems with packet loss," *IEEE Infocom AoI Workshop*, 2019.
- [19] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," *arXiv preprint arXiv:1903.05356*, 2019.
- [20] A. Maatouk, S. Kriouile, M. Assad, and A. Ephremides, "On the optimality of the whittle's index policy for minimizing the age of information," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 1263–1277, 2020.
- [21] V. Tripathi and E. Modiano, "A whittle index approach to minimizing functions of age of information," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1160–1167, IEEE, 2019.
- [22] P. Whittle, "Restless bandits: Activity allocation in a changing world," *Journal of applied probability*, vol. 25, no. A, pp. 287–298, 1988.
- [23] J. C. Gittins, K. D. Glazebrook, R. Weber, and R. Weber, *Multi-armed bandit allocation indices*, vol. 25. Wiley Online Library, 1989.
- [24] V. Tripathi, L. Ballotta, L. Carlone, and E. Modiano, "Computation and communication co-design for real-time monitoring and control in multi-agent systems," in *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad hoc, and Wireless Networks (WiOpt)*, pp. 1–8, IEEE, 2021.

APPENDIX

A. Proof of Theorem 1

Consider the decoupled problem described in Section IV. Let $u(t)$ be an indicator variable that denotes whether the arm is pulled or not at time t . Under a scheduling policy π that specifies the value of $u(t)$ for all instants of time, the average cost is given by

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \left[f(A^\pi(t)) + C u^\pi(t) \right]. \quad (21)$$

We want to find a policy that minimizes this cost over the space of all policies. Let $S : \mathbb{Z}^+ \rightarrow \mathbb{R}$ denote the differential cost-to-go function for this problem, let $u : \mathbb{Z}^+ \rightarrow \{1, 0\}$ be the stationary optimal policy and let λ denote the optimal cost. Then, the Bellman equations are given by

$$S(h) = f(h) + \min_{u(h) \in \{1, 0\}} \{C, S(h+1)\} - \lambda, \forall h \in \mathbb{Z}^+. \quad (22)$$

Without loss of generality we set $S(1) = 0$. Assume that the optimal policy has a threshold structure, i.e. there exists H such that it is optimal to pull the arm ($u(h) = 1$) for all states $h \geq H$ and let it rest otherwise ($u(h) = 0$). If this the case, then the Bellman equations reduce to

$$S(h) = f(h) + C - \lambda, \forall h \geq H. \quad (23)$$

Using the monotonicity of $f(\cdot)$, we conclude that $S(h+1) \geq S(h)$, $\forall h \geq H$. We will use this fact later. For the state $H-1$, we get

$$\begin{aligned} S(H-1) &= f(H-1) - \lambda + S(H) \\ &= f(H-1) - \lambda + f(H) - \lambda + C. \end{aligned} \quad (24)$$

Repeating this k times, we get

$$S(H-k) = \sum_{j=0}^k f(H-j) - (k+1)\lambda + C, \quad (25)$$

for all k in $\{1, \dots, H-1\}$. Observe that since we set $S(1) = 0$, we get

$$\lambda = \frac{\sum_{j=1}^H f(j) + C}{H}, \quad (26)$$

by putting $k = H-1$ in (25). Now assume that H further satisfies the relation given in Theorem 1, i.e.

$$f(H) \leq \frac{\sum_{j=1}^H f(j) + C}{H} \leq f(H+1). \quad (27)$$

Using (26), we can simplify (27) as

$$f(H) \leq \lambda \leq f(H+1). \quad (28)$$

Adding $C - \lambda$ to every term above, we get

$$\begin{aligned} f(H) + C - \lambda &\leq C \leq f(H+1) + C - \lambda \\ \implies S(H) &\leq C \leq S(H+1). \end{aligned} \quad (29)$$

Observe that we assumed $f(\cdot)$ to be non-decreasing. This combined with (28) and the Bellman equations (23) and (25) ensures that $S(\cdot)$ is also non-decreasing. Thus, if there exists a state H that satisfies (27), then the threshold policy with threshold H satisfies the Bellman equations and is hence optimal.

The one thing that remains to be shown is the case in which we cannot find some H that satisfies (27). Consider the function $W : \mathbb{Z}^+ \rightarrow \mathbb{R}$ given by

$$W(h) = hf(h) - \sum_{j=1}^h f(j). \quad (30)$$

Observe that $W(h+1) - W(h) = h(f(h+1) - f(h)) \geq 0$ since $f(\cdot)$ is non-decreasing. Thus, $W(\cdot)$ is also non-decreasing. Also, by definition, $W(1) = 0$, while we had assumed that $C > 0$. Thus, $W(1) < C$. Now, if there exists some $h > 1$ such that $W(h) \geq C$, then we know that there also exists some H such that $W(H) \leq C \leq W(H+1)$ using monotonicity of $W(\cdot)$. Observe that this implies that there exists some H satisfying

$$Hf(H) - \sum_{j=1}^H f(j) \leq C \leq (H+1)f(H+1) - \sum_{j=1}^{H+1} f(j).$$

Rearranging and dividing by H , we get back (27). Thus, if there exists no H satisfying (27), then $W(h) < C, \forall h$.

Since $W(\cdot)$ is a bounded monotone sequence, it converges to a finite value. It is easy to see that this implies that $f(\cdot)$ is also bounded and hence converges. We set $\lambda = \lim_{h \rightarrow \infty} f(h)$ and the cost-to-go function $S(h)$ to be

$$S(h) = \sum_{j=h}^{\infty} (f(j) - \lambda) + C. \quad (31)$$

Clearly, $S(h)$ satisfies the recurrence relation

$$S(h) = f(h) - \lambda + S(h+1), \forall h. \quad (32)$$

By the monotonicity of $f(\cdot)$, we know that $f(h) \leq \lambda, \forall h$. Thus, using (31) we conclude that $S(h) \leq C, \forall h$. This implies that $S(\cdot)$ satisfies the Bellman equations, with the optimal policy being to never activate the arm. This completes our proof.

B. Proof of Theorem 2

For $C = 0$, it is obvious that the optimal policy is to always activate the arm since there is no charge for activating it and the cost function is monotone and positive. For larger values of C , consider the function $W : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ given by

$$W(h) = hf(h) - \sum_{j=1}^h f(j). \quad (33)$$

Observe that since $f(\cdot)$ is non-decreasing, $W(\cdot)$ is also non-decreasing. This is because $W(h+1) - W(h) = h(f(h+1) - f(h)) \geq 0, \forall h$ since $f(\cdot)$ is non-decreasing. Also, by definition, $W(1) = 0$, while we had assumed that $C > 0$. Thus, $W(1) < C$. Now, if there exists some $h > 1$ such that $W(h) \geq C$, then we know that there also exists some H such that $W(H) \leq C \leq W(H+1)$ using monotonicity of $W(\cdot)$. Observe that this implies that there exists some H satisfying

$$Hf(H) - \sum_{j=1}^H f(j) \leq C \leq (H+1)f(H+1) - \sum_{j=1}^{H+1} f(j). \quad (34)$$

Rearranging and dividing by H , we get back (27).

Using this, we can relate the optimal threshold values to values of activation charge. Let C be such that it lies in the interval $[W(h), W(h+1))$, then the optimal policy is of threshold type with the threshold at h . Observe that if W is strictly increasing then there can only be one such interval in which C can lie. If $W(\cdot)$ is non-decreasing, then there could be multiple such intervals in which C could lie. In this case, we choose the smallest h such that the condition holds.

The monotonicity of $W(\cdot)$ ensures that the threshold value is also monotone non-decreasing with increasing values of C . When $W(h) < C, \forall h$, we choose h to be ∞ , as done in Appendix A. This completes the proof of *indexability* for the decoupled problem. Observe that $C = W(h+1)$ is the minimum value of the activation charge that makes both actions equally desirable in state h . This gives us the expression for the Whittle index.

C. Proof of Theorem 4

We look at the optimal cyclical policy and analyze its properties. If there are multiple such cycles, we consider the cycle with the shortest length. We denote the length of the cycle by T , points on the cycle to be $\mathbf{x}_1, \dots, \mathbf{x}_T$, and the average cost of this cycle to be C^* . The point \mathbf{x}_i is an age vector in \mathbb{Z}^{+N} , where $x_i^{(j)}$ represents the age of the j^{th} source. Let the corresponding scheduling decisions be d_1, \dots, d_T . This implies that for age vector \mathbf{x}_k , taking action d_k leads

to the age vector \mathbf{x}_{k+1} , where the subscripts cycle back to $1, 2, \dots$ after T . Assume that there exists some pair of states in this cycle that violate the strong-switch-type property. If not, then our claim that the cycle satisfies the strong-switch-type property is true.

Without loss of generality, we assume that the pair of states that violates strong-switch is given by \mathbf{x}_1 and \mathbf{x}_k for some $k \in \{2, \dots, T\}$. This is because the cyclical policy is same up to cyclical permutations, so we can always ensure that one member of the violating pairs is at the front of the cycle. Also without loss of generality, we assume that $d_1 = 1$ and $d_k = 2$, since we can always relabel the sources. Observe that d_1 and d_k cannot be the same since they violate the strong-switch property. In fact, we know that $x_k^{(j)} \leq x_1^{(j)}, \forall j \neq 1$ and $x_k^{(1)} \geq x_1^{(1)}$. If the strong-switch property was satisfied, d_k must have been 1 since d_1 is 1.

We now construct two new cyclical policies out of which at least one has a better cost or the same cost but a smaller length compared to the original optimal policy. This contradicts our original assumption that the cycle we had started with was the shortest policy with the lowest average cost. Starting with the state \mathbf{x}_1 , we take the action d_1 , following the original cycle up to \mathbf{x}_k . At \mathbf{x}_k , instead of taking action d_k , we take the action d_1 leading to the state \mathbf{y}_{k+1} . Observe that $\mathbf{y}_{k+1} \leq \mathbf{x}_2$, where the inequality is element-wise. Since $d_1 = 1$, we schedule source 1 at both \mathbf{x}_1 and \mathbf{x}_k , which guarantees that its age goes to 1. Thus, $x_2^{(1)} = y_{k+1}^{(1)} = 1$. Also, since $x_k^{(j)} + 1 \leq x_1^{(j)} + 1, \forall j \neq 1$ and none of the other sources are scheduled, so $y_{k+1}^{(j)} \leq x_2^{(j)}, \forall j \neq 1$. Together, this implies $\mathbf{y}_{k+1} \leq \mathbf{x}_2$.

Now, we follow the original cycle starting from d_2, \dots, d_T . Action d_2 at state \mathbf{y}_{k+1} leads to state \mathbf{y}_{k+2} and so on, up to action d_T at state \mathbf{y}_{k+T-1} . Since the channels are reliable and $\mathbf{y}_{k+1} \leq \mathbf{x}_2$, it is easy to see that $\mathbf{y}_{k+i} \leq \mathbf{x}_{i+1}, \forall i \in \{1, \dots, T-1\}$.

Also, observe that starting at \mathbf{x}_k , we have repeated an entire period of the original cycle, i.e. d_1, \dots, d_T . Every source gets activated at least once during the original cycle, otherwise, its age goes to infinity, and we might as well remove it from the system. Starting at any age vector and following the actions d_1, \dots, d_T in sequence ensures that the state reached after these T steps equals \mathbf{x}_1 . Thus, the actions $\{d_1, \dots, d_{k-1}, d_1, \dots, d_T\}$ and the age vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_k, \mathbf{y}_{k+1}, \dots, \mathbf{y}_{k+T-1}\}$ form a cycle of length $k+T-1$. Here $\mathbf{y}_i = \mathbf{x}_i, \forall i \in \{1, \dots, k\}$ and $\mathbf{y}_{k+i} \leq \mathbf{x}_{i+1}, \forall i \geq 1$. We denote the average cost of this cycle by C_1 .

Now, we perform a cyclic permutation of the original optimal policy to get a new optimal policy with the actions $\{d_k, \dots, d_T, d_1, \dots, d_{k-1}\}$ and the corresponding states $\{\mathbf{x}_k, \dots, \mathbf{x}_T, \mathbf{x}_1, \dots, \mathbf{x}_{k-1}\}$. We repeat the process of constructing a new cyclical policy of length $2T - k + 1$ as done above, but using the new cyclic permutation of the optimal policy. That is, instead of choosing action d_1 at \mathbf{x}_k , we choose action d_k at \mathbf{x}_1 .

This new cyclical policy consists of actions $\{d_k, \dots, d_T, d_k, \dots, d_T, d_1, \dots, d_{k-1}\}$ and the corresponding

age vectors $\{z_1, \dots, z_{2T-k+1}\}$, forming a cycle of length $2T - k + 1$. Using exactly the same argument as earlier, it is easy to see that $z_j = x_{j+k-1}, \forall j \in \{1, \dots, T - k + 1\}$ and $z_j \leq x_j, \forall j \in \{T - k + 2, \dots, 2T - k + 1\}$. We denote the average cost of this cycle by C_2 .

We know that the cost of the optimal policy C^* is minimum over the space of all policies, and hence less than or equal to cost of the first cyclical policy that we created C_1 . Thus,

$$\begin{aligned} C^* &\leq C_1 \\ \Rightarrow \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}) &\leq \frac{1}{k+T-1} \sum_{t=1}^{k+T-1} \sum_{j=1}^N f_j(y_t^{(j)}) \\ &\leq \frac{1}{k+T-1} \left(\sum_{t=1}^k \sum_{j=1}^N f_j(x_t^{(j)}) + \sum_{t=k+1}^{k+T-1} \sum_{j=1}^N f_j(y_t^{(j)}) \right) \\ &\leq \frac{1}{k+T-1} \left(\sum_{t=1}^k \sum_{j=1}^N f_j(x_t^{(j)}) + \sum_{t=2}^T \sum_{j=1}^N f_j(x_t^{(j)}) \right) \end{aligned}$$

Simplifying this inequality, we get

$$\frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}) \leq \frac{1}{k-1} \sum_{t=2}^k \sum_{j=1}^N f_j(x_t^{(j)}). \quad (35)$$

Similarly, the average cost of the optimal cycle $*$ is also less than or equal to the average cost of the second cycle C_2 . Thus,

$$\begin{aligned} C^* &\leq C_2 \\ \Rightarrow \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}) &\leq \frac{1}{2T-k+1} \sum_{t=1}^{2T-k+1} \sum_{j=1}^N f_j(z_t^{(j)}) \\ &\leq \frac{1}{2T-k+1} \left(\sum_{t=k}^T \sum_{j=1}^N f_j(x_t^{(j)}) + \sum_{j=1}^N f_1(x_1^{(j)}) + \sum_{t=T-k+3}^{2T-k+1} \sum_{j=1}^N f_j(z_t^{(j)}) \right) \\ &\leq \frac{1}{2T-k+1} \left(\sum_{t=k+1}^T \sum_{j=1}^N f_j(x_t^{(j)}) + \sum_{j=1}^N f_1(x_1^{(j)}) + \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}) \right) \end{aligned}$$

Simplifying this inequality, we get

$$\frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}) \leq \frac{1}{T-k+1} \left(\sum_{t=k+1}^T \sum_{j=1}^N f_j(x_t^{(j)}) + \sum_{j=1}^N f_1(x_1^{(j)}) \right).$$

Rearranging and simplifying again, we get

$$\frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}) \geq \frac{1}{k-1} \sum_{t=2}^k \sum_{j=1}^N f_j(x_t^{(j)}). \quad (36)$$

From the analysis above, we observe that (35) and (36) must hold simultaneously. However, if that's the case then the

inequalities cannot be strict. Also, observe that the cyclical policy given by actions $\{d_2, \dots, d_k\}$ has average cost C_3 that satisfies

$$C_3 \leq \frac{1}{k-1} \sum_{t=2}^k \sum_{j=1}^N f_j(x_t^{(j)}).$$

This is because starting at state x_2 and following the policy we end up at state x_k , where using the exact same argument as earlier, taking action d_k leads us to a state y_{k+1} such that $y_{k+1} \leq x_2$. The upper bound follows directly. Also, since (35) is tight, we get that

$$C_3 \leq \frac{1}{T} \sum_{t=1}^T \sum_{j=1}^N f_j(x_t^{(j)}).$$

This is a contradiction, since if the above inequality is strict, our original policy is no longer optimal and if the inequality is tight, we have a smaller length cycle with the same cost, which still contradicts our original assumption that we started with an optimal cost cycle with minimum length.

D. Proof of Theorem 5

We have shown that the points on the optimal cycle satisfy the strong-switch-type property. We need to show that we can assign actions to states that are not on the optimal cycle while maintaining the strong-switch property, for $N \leq 3$.

This can be done in an iterative manner. Consider the set of points in the state-space that have been assigned an action, and which satisfy the strong-switch property to be D . Let $x \notin D$, be a new point that we want to assign an action to. There are three possible scenarios - 1) there exists $y \in D$ such that the strong-switch-type property implies a unique action to be taken at x , 2) there exists no such $y \in D$ and so an arbitrary action can be chosen at x , and 3) there exist multiple such points in D , which suggest different actions to be taken at x .

Clearly, for scenarios 1 and 2 above, we can assign an action to the point x , increase our set to $D \cup \{x\}$ and repeat the procedure for a new point. We claim that if $N \leq 3$ then scenario 3 never occurs. This is sufficient to prove that we can extend the strong-switch-type property over the entire state-space.

To prove that scenario 3 doesn't happen, we start by assuming the contrary. Let y_1 and $y_2 \in D$ and without loss of generality, assume that the action taken at y_1 is 1 and the action taken at y_2 is 2. Also, to satisfy our assumption of scenario 3 for x , we require that $x^{(1)} \geq y_1^{(1)}$, $x^{(2)} \geq y_2^{(2)}$, $x^{(j)} \leq y_1^{(j)}, \forall j \neq 1$ and $x^{(k)} \leq y_1^{(k)}, \forall k \neq 2$. For these inequalities to be feasible simultaneously, we need $y_2^{(1)} \geq y_1^{(1)}$ and $y_1^{(2)} \geq y_2^{(2)}$.

Now, if there are only two sources, i.e. $N = 2$, then the fact that $y_2^{(1)} \geq y_1^{(1)}$ and $y_1^{(2)} \geq y_2^{(2)}$ together with the assumption that the action taken at y_1 is 1 and the action taken at y_2 is 2, we get that y_1 and y_2 violate the strong-switch property, despite being in the set D . This is a contradiction and completes our proof.

Similarly, consider the setting with three sources ($N=3$). Now, there are two possibilities - either $y_1^{(3)} \leq y_2^{(3)}$ or $y_1^{(3)} > y_2^{(3)}$. If $y_1^{(3)} \leq y_2^{(3)}$, then using the fact that $y_1^{(1)} \leq y_2^{(1)}$ and $y_1^{(2)} \geq y_2^{(2)}$, the strong-switch-type property implies that the action taken at \mathbf{y}_1 must be 2. However, we assumed that the action taken at state \mathbf{y}_1 is 1. Thus, this violates the strong-switch-type property. Similarly, if $y_1^{(3)} > y_2^{(3)}$, then using the fact that $y_1^{(1)} \leq y_2^{(1)}$ and $y_1^{(2)} \geq y_2^{(2)}$, the strong-switch-type property implies that the action taken at \mathbf{y}_2 must be 1. This again violates our assumption that \mathbf{y}_1 and \mathbf{y}_2 satisfy the strong-switch-type property.

Thus, we have proved that for $N \leq 3$, if the optimal cycle is strong-switch-type then we can find a stationary optimal policy that is strong-switch-type over the entire state-space.

E. Proof of Theorem 6

We use an inductive argument to prove this result. Assume that for a scheduling setup with $N - 1$ sources, every strong-switch type policy can also be written as an index policy. Now, consider a functions of age setup with N sources and reliable channels. Using Theorem 5 we know that there exists an optimal policy that is strong-switch-type. Let this policy be $\pi : \mathbb{Z}^+ \rightarrow \{1, \dots, N\}$.

Let x_i denote the age of the i^{th} source when the current state is \mathbf{x} and let \mathbf{x}_{-i} denote the vector comprising of ages of all sources except i . Consider the minimum age x_1 at source 1 such that $\pi(x_1, \mathbf{x}_{-1}) = 1$, for any fixed \mathbf{x}_{-1} . That is, for any fixed value of ages for all other sources, consider the age at source 1 for which the optimal policy schedules the first source. This value of x_1 may depend on \mathbf{x}_{-1} , so we denote it by $x_{\text{th}}(\mathbf{x}_{-1})$. Observe that for all values of x_1 such that $x_1 \geq x_{\text{th}}(\mathbf{x}_{-1})$, the strong-switch-type property implies that $\pi(x_1, \mathbf{x}_{-1}) = 1$. In other words, $x_{\text{th}}(\mathbf{x}_{-1})$ acts like a threshold value such that for all values of age at source 1 above it, the optimal policy schedules the first source. If no such threshold exists, we let $x_{\text{th}}(\mathbf{x}_{-1}) \rightarrow \infty$.

We append the state space of the first arm by zero, i.e. let the state space of source 1 be $\mathbb{Z}_0 = \mathbb{Z}^+ \cup \{0\}$. Zero is the minimum age that this source can have and without loss of generality, we can set $f_1(0) = 0$. If at any time-slot the age of this source is zero, we let it increases to one in the next time-slot. Scheduling this source when its age is zero gives us no benefit, as the age increases by one no matter what our scheduling decision is. We extend the policy π over this new state space as follows. Let $\pi' : \mathbb{Z}_0 \times \mathbb{Z}^{+N-1} \rightarrow \{1, \dots, N\}$ be a mapping that satisfies

- $\pi'(\mathbf{x}) = \pi(\mathbf{x})$, $\forall \mathbf{x}$ s.t. $x_1 \neq 0$.
- if $x_{\text{th}}(\mathbf{x}_{-1}) > 1$, $\pi'(0, \mathbf{x}_{-1}) = \pi(x_{\text{th}}(\mathbf{x}_{-1}) - 1, \mathbf{x}_{-1})$
- if $x_{\text{th}}(\mathbf{x}_{-1}) = 1$, $\pi'(0, \mathbf{x}_{-1}) = \pi_{N-1}(\mathbf{x}_{-1})$,

where $\pi_{N-1}(\mathbf{x}_{-1})$ is an optimal strong-switch-type policy for the functions of age problem with just the sources $2, \dots, N$. It is easy to see that this new extended policy still satisfies the strong-switch-type property and is still optimal for the original problem with all N sources over the extended state space.

Now, we project this new optimal policy π' on to \mathbb{Z}^{+N-1} to get a new policy $\pi'' : \mathbb{Z}^{+N-1} \rightarrow \{2, \dots, N\}$ such that

$\pi''(\mathbf{x}_{-1}) = \pi'(x_{\text{th}}(\mathbf{x}_{-1}) - 1, \mathbf{x}_{-1})$. This is well defined since $x_{\text{th}}(\mathbf{x}_{-1}) \geq 1$ and $\pi'(x_{\text{th}}(\mathbf{x}_{-1}) - 1, \mathbf{x}_{-1}) \in \{2, \dots, N\}$, by construction. Also, π'' is strong-switch-type by construction, since it is a projection of a strong-switch-type policy onto a lower dimensional space. If not, then π' would also violate the strong-switch-type property.

Now, using our induction assumption, we can find index functions such that

$$\pi''(x_2, \dots, x_N) = \arg \max_{2 \leq i \leq N} \left\{ F_i(x_i) \right\}$$

for all \mathbf{x} , where $F_i : \mathbb{Z}^+ \rightarrow \mathbb{R}$ are monotonically non-decreasing functions for all i .

We partition the $N - 1$ dimensional state space of policy π'' into a countable number of sets. Let

$$S_k \triangleq \{ \mathbf{x} : \mathbf{x} \in \mathbb{Z}^{+N-1}, x_{\text{th}}(\mathbf{x}) = k \}, \forall k \in \mathbb{Z}^+.$$

Then, $\mathbb{Z}^{+N-1} = \bigcup_{k=1}^{\infty} S_k$ and $S_k \cap S_j = \{\emptyset\}$, $\forall k, j$. Consider $\mathbf{x} \in S_j$ and $\mathbf{y} \in S_k$ such that $k > j$. Then,

$$\max_{2 \leq i \leq N} \left\{ F_i(x_i) \right\} \leq \max_{2 \leq i \leq N} \left\{ F_i(y_i) \right\}. \quad (37)$$

Suppose the opposite is true, i.e. $\max_{2 \leq i \leq N} \left\{ F_i(x_i) \right\} > \max_{2 \leq i \leq N} \left\{ F_i(y_i) \right\}$. Let $m = \arg \max_{2 \leq i \leq N} \left\{ F_i(x_i) \right\}$. Clearly, for the opposite of (37) to hold we need $x_m > y_m$. If we define \mathbf{z} such that $z_i \triangleq \max\{x_i, y_i\}$, $\forall i \in 2, \dots, N$, then using the index property of $\pi''(\cdot)$ we get

$$\pi''(\mathbf{z}) = m.$$

Also, $x_{\text{th}}(\mathbf{z}) \geq k$. If not, then since $\mathbf{z} \geq \mathbf{y}$, the strong-switch property implies $\pi'(x_{\text{th}}(\mathbf{z}), \mathbf{y}) = 1$, where $x_{\text{th}}(\mathbf{z}) < k$. This violates our assumption that $\mathbf{y} \in S_k$.

Now, observe that $\pi'(x_{\text{th}}(\mathbf{z}) - 1, \mathbf{z}) = m$, since $\pi''(\mathbf{z}) = m$. Also $(j, x_2, \dots, x_m, \dots, x_N) \leq (x_{\text{th}}(\mathbf{z}) - 1, z_2, \dots, z_m, \dots, z_N)$ where the inequality holds element-wise. This is because $z_i = \max\{x_i, y_i\}$, and $j \leq k - 1 \leq x_{\text{th}}(\mathbf{z}) - 1$, and $x_m > y_m$. Thus, using the strong-switch-type property of $\pi'(\cdot)$, we get

$$\pi'(j, \mathbf{x}) = m.$$

This contradicts our initial assumption that $\mathbf{x} \in S_j$, since that would imply $\pi'(j, \mathbf{x}) = 1$. Thus, we conclude that (37) must be satisfied.

We now construct a monotone function based on the above discussion. Let

$$F_1(j) \triangleq \sup_{\mathbf{z} \in S_j} \max_{2 \leq i \leq N} \left\{ F_i(z_i) \right\}, \forall j \in \mathbb{Z}^+. \quad (38)$$

Clearly, since the condition (37) is satisfied, the function $F_1(\cdot)$ is monotone. Also, let

$$\pi'''(\mathbf{x}) = \arg \max_{1 \leq i \leq N} \left\{ F_i(x_i) \right\}, \forall \mathbf{x} \in \mathbb{Z}^{+N},$$

where we break ties in lexicographic order. Then π''' is the same as our original policy π . This is because for every state $\mathbf{x} \in$

S_j , the construction of F_1 forces us to schedule source 1 for values of $x_1 \geq j$ and not schedule source 1 for values below j . This holds for all values of j , which means we replicate the original scheduling policy $\pi(\cdot)$. Thus, if we assume strong-switch-type policies can be written as index policies for a problem with $N - 1$ sources, we can also prove the same fact for N sources.

It is trivial to see that strong-switch-type policies and index policies are equivalent for the single source decoupled problem. This is because strong-switch-type policies and index policies both correspond to monotone threshold policies for the decoupled problem. Hence, using the principle of induction, we have the required result.

F. Proof of Theorem 3

Using Corollary 1, we know that there exists some index policy which is optimal. We observe that for $N = 2$ index policies have a specific structure.

Let $F_1(\cdot)$ and $F_2(\cdot)$ represent the index functions for the optimal index policy. We set the ages of the two sources to $(1, 1)$ at time $t = 1$ and assume that the optimal index functions are such that $F_1(1) \geq F_2(1)$. Then, the policy schedules source 1 at time $t = 1$. The new state at time $t = 2$ is given by $(1, 2)$. Again, assume that $F_1(1) \geq F_2(2)$. Then, the policy schedules source 1 at time $t = 2$. The new state at time $t = 3$ is given by $(1, 3)$. We keep repeating this process until we reach state $(1, k)$ at time $t = k$ for which $F_1(1) < F_2(k)$. The policy then schedules source 2 and reaches state $(2, 1)$ at time $k + 1$. Now, since we assumed that $F_1(1) \geq F_2(1)$, then using monotonicity we get $F_1(2) \geq F_2(1)$. Thus, the policy schedules source 1 again and we reach state $(1, 2)$ at time $t = k + 2$.

From the above discussion, we see that any index policy for $N = 2$ has a cyclic form and the cycle consists one of the sources being scheduled repeatedly followed by the second source once. To find the best index policy, which is also the best policy overall, we just need to find the best policy with this specific structure.

Without loss of generality, assume that an optimal cyclical policy is given by scheduling source 1 k times followed by source 2 once, and repeating this sequence of actions. Now, consider two cases.

1) *Case 1:* ($k > 1$) We compare the cost of the optimal cycle with a cycle that schedules source 1 $k - 1$ times followed by source 2 once.

$$\frac{\sum_{j=1}^k f_2(j) + (k - 1)f_1(1) + f(2)}{k} \geq \frac{\sum_{j=1}^{k+1} f_2(j) + kf_1(1) + f(2)}{k + 1}$$

Simplifying, we get

$$f_1(2) - f_1(1) \geq kf_2(k + 1) - \sum_{j=1}^k f_2(j),$$

i.e. $W_1(1) \geq W_2(k)$. The Whittle index policy follows the optimal policy till the state $(1, k)$.

We then compare the cost of the optimal cycle with a cycle that schedules source 1 $k + 1$ times followed by source 2 once.

$$\frac{\sum_{j=1}^{k+2} f_2(j) + (k + 1)f_1(1) + f(2)}{k + 2} > \frac{\sum_{j=1}^{k+1} f_2(j) + kf_1(1) + f(2)}{k + 1}$$

Simplifying, we get

$$(k + 1)f_2(k + 2) - \sum_{j=1}^{k+1} f_2(j) > f_1(2) - f_1(1),$$

i.e. $W_2(k + 1) > W_1(1)$.

Together, this implies that the Whittle Index policy must also schedule source 1 k times followed by source 2 once, and repeat this sequence of actions. Hence, the Whittle index policy is optimal.

2) *Case 2:* ($k = 1$) We compare the optimal policy with a cycle that schedules source 1 twice and source 2 once. Then, we get

$$\frac{2f_1(1) + f_1(2) + f_2(1) + f_2(2) + f_2(3)}{3} > \frac{f_1(1) + f_2(1) + f_1(2) + f_2(2)}{2}$$

Simplifying, we get

$$2f_2(3) - f_2(1) - f_2(2) > f_1(2) - f_1(1),$$

i.e. $W_2(2) > W_1(1)$. Using a symmetrical argument, it is easy to see that $W_1(2) > W_2(1)$. Thus, the Whittle Index policy also schedules each source exactly once, and repeats this sequence of actions.

Combining the two cases, we conclude that for $N = 2$ and reliable channels, the Whittle Index policy is exactly optimal.

G. Proof of Theorem 7

Let $S : \mathbb{Z}^+ \rightarrow \mathbb{R}$ denote the differential cost-to-go function for this problem, let $u : \mathbb{Z}^+ \rightarrow \{1, 0\}$ be the stationary optimal policy and let λ denote the optimal cost. Then, the Bellman equations are given by

$$S(h) = f(h) + \min_{u(h) \in \{1, 0\}} \{C + (1 - p)S(h + 1), S(h + 1)\} - \lambda, \forall h \in \mathbb{Z}^+. \quad (39)$$

Without loss of generality we set $S(1) = 0$. Assume that the optimal policy has a threshold structure, i.e. there exists H such that it is optimal to pull the arm ($u(h) = 1$) for all states $h \geq H$ and let it rest otherwise ($u(h) = 0$). If this the case, then the Bellman equations for values above the threshold H reduce to

$$S(h) = f(h) + C + (1 - p)S(h + 1) - \lambda, \forall h \geq H. \quad (40)$$

Solving this recursion and assuming $\lim_{h \rightarrow \infty} (1 - p)^h S(h) = 0$, we get

$$S(H + j) = \sum_{k=j}^{\infty} f(k + H)(1 - p)^{k-j} + \frac{C - \lambda}{p}, \forall j \geq 0. \quad (41)$$

Since $f(\cdot)$ is non-decreasing, it is easy to see that $S(h)$ is also non-decreasing for all values of h above the threshold H , using (41). We will use this fact later. Now, observe that

$$\begin{aligned} \lim_{h \rightarrow \infty} (1-p)^h S(h) &= \lim_{h \rightarrow \infty} \sum_{j=h}^{\infty} f_i(j)(1-p_i)^j \\ &+ \lim_{h \rightarrow \infty} \frac{C-\lambda}{h} (1-p_i)^h, \forall h \geq H. \end{aligned} \quad (42)$$

By the bounded cost assumption, the first term is the limit of the partial sums of a convergent series, thus it goes to zero. The second term also goes to zero since $p < 1$ and λ is finite, again using the bounded cost assumption. This confirms that our assumption $\lim_{h \rightarrow \infty} (1-p)^h S(h) = 0$ was indeed correct.

For $h = H - 1$, the Bellman equation is given by

$$\begin{aligned} S(H-1) &= f(H-1) - \lambda + S(H) \\ &= f(H-1) - \lambda + \sum_{k=0}^{\infty} f(k+H)(1-p)^k + \frac{C-\lambda}{p}. \end{aligned} \quad (43)$$

Repeating this k times, we get

$$\begin{aligned} S(H-k) &= \sum_{j=H-k}^{H-1} (f(j) - \lambda) + \sum_{j=0}^{\infty} f(j+H)(1-p)^j \\ &+ \frac{C-\lambda}{p}, \forall k \in \{1, \dots, H-1\}. \end{aligned} \quad (44)$$

Now, putting $k = H - 1$ in the above equation and using the fact that $S(1) = 0$, we get

$$\lambda = \frac{p(\sum_{j=1}^H f(j) + \sum_{k=1}^{\infty} f(k+H)(1-p)^k) + C}{1+p(H-1)}. \quad (45)$$

If we further assume that the threshold value H satisfies the condition (11) given in Theorem 7, then we get that

$$\begin{aligned} p^2(H-1) \left(\sum_{k=H}^{\infty} f(k)(1-p)^{k-H} \right) - p \left(\sum_{j=1}^{H-1} f(j) \right) \\ \leq C \\ \leq p^2 H \left(\sum_{k=H+1}^{\infty} f(k)(1-p)^{k-H-1} \right) - p \left(\sum_{j=1}^H f(j) \right). \end{aligned} \quad (46)$$

Rearranging terms, dividing by $1+p(H-1)$ and using the expression for λ from (45), we get

$$\begin{aligned} p \left(\sum_{k=0}^{\infty} f(H+k)(1-p)^k \right) &\leq \lambda \\ &\leq p \left(\sum_{k=1}^{\infty} f(H+k)(1-p)^{k-1} \right). \end{aligned} \quad (47)$$

Simplifying the inequalities in (47), the expression for λ from (45) and the Bellman solutions (41), we get

$$S(H) \leq \frac{C}{p} \leq S(H+1). \quad (48)$$

Using (44), we note that for $h < H$, $S(h) - S(h-1) = \lambda - f(h-1)$. Also, using the monotonicity of $f(\cdot)$ and (47), we get

$$\begin{aligned} \lambda &\geq p \left(\sum_{k=0}^{\infty} f(H+k)(1-p)^k \right) \\ &\geq p \left(\sum_{k=0}^{\infty} f(H)(1-p)^k \right) \\ &\geq f(H) \\ &\geq f(h), \forall h < H. \end{aligned} \quad (49)$$

Thus, $S(h) - S(h-1) \geq 0, \forall h$ since we already established monotonicity for $h \geq H$. Since $S(\cdot)$ is non-decreasing, (48) implies that

$$\begin{aligned} S(h) &\leq C + (1-p)S(h), \forall h \leq H, \text{ and} \\ S(h) &\geq C + (1-p)S(h), \forall h > H. \end{aligned} \quad (50)$$

Thus, if we find an H that satisfies (11), the threshold policy using H as a threshold satisfies the Bellman equations and is optimal.

The one thing that remains to be shown is the case in which we cannot find some H that satisfies (27). As done earlier, we define a function $W : \mathbb{Z}^+ \rightarrow \mathbb{R}$ given by

$$W(h) = p^2(h-1) \left(\sum_{k=h}^{\infty} f(k)(1-p)^{k-h} \right) - p \left(\sum_{j=1}^{h-1} f(j) \right). \quad (51)$$

Observe that

$$\begin{aligned} W(h+1) - W(h) &= \\ p^2 h \left[\sum_{k=0}^{\infty} (f(h+1+k) - f(h+k))(1-p)^k \right] \\ &+ p^2 \left[\sum_{k=0}^{\infty} f(h+k)(1-p)^k \right] - p f(h) \\ &\geq 0, \forall h \end{aligned} \quad (52)$$

since $f(\cdot)$ is non-decreasing. Thus, $W(\cdot)$ is also non-decreasing. Also, putting $h = 1$ in the definition of $W(h)$ we get $W(1) = 0$, while we had assumed that $C > 0$. Thus, $W(1) < C$. Now, if there exists some $h > 1$ such that $W(h) \geq C$, then we know that there also exists some H such that $W(H) \leq C \leq W(H+1)$ using monotonicity of $W(\cdot)$. Observe that this implies that there exists some H satisfying (11) and hence the threshold policy is optimal. If there exists no H satisfying (11), then $W(h) < C, \forall h$.

Since $W(\cdot)$ is a bounded monotone sequence, it converges to a finite value. It is easy to see that this implies that $f(\cdot)$ is also bounded and hence converges. We set $\lambda = \lim_{h \rightarrow \infty} f(h)$ and the cost-to-go function $S(h)$ to be

$$S(h) = \sum_{j=h}^{\infty} (f(j) - \lambda) + C. \quad (53)$$

Clearly, $S(h)$ satisfies the recurrence relation

$$S(h) = f(h) - \lambda + S(h+1), \forall h. \quad (54)$$

By the monotonicity of $f(\cdot)$, we know that $f(h) \leq \lambda, \forall h$. Thus, using (53) we conclude that $S(h) \leq C, \forall h$. This implies that $S(\cdot)$ satisfies the Bellman equations, with the optimal policy being to never activate the arm. This completes our proof.

H. Proof of Theorem 8

This proof is very similar to the indexability proof for the reliable channels case. For $C = 0$, it is obvious that the optimal policy is to always activate the arm since there is no charge for activating it and the cost function is monotone and positive. For larger values of C , consider the function $W : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ given by

$$W(h) = p^2(h-1) \left(\sum_{k=h}^{\infty} f(k)(1-p)^{k-h} \right) - p \left(\sum_{j=1}^{h-1} f(j) \right).$$

Observe that since $f(\cdot)$ is non-decreasing, $W(\cdot)$ is also non-decreasing, as discussed in Appendix G. Also, by definition, $W(1) = 0$, while we had assumed that $C > 0$. Thus, $W(1) < C$. Now, if there exists some $h > 1$ such that $W(h) \geq C$, then we know that there also exists some H such that $W(H) \leq C \leq W(H+1)$ using monotonicity of $W(\cdot)$. Observe that this implies that there exists some H satisfying (11).

Using this, we can relate the optimal threshold values to values of activation charge. Let C be such that it lies in the interval $[W(h), W(h+1))$, then the optimal policy is of threshold type with the threshold at h . Observe that if W is strictly increasing then there can only be one such interval in which C can lie. If $W(\cdot)$ is non-decreasing, then there could be multiple such intervals in which C could lie. In this case, we choose the smallest h such that the condition holds.

The monotonicity of $W(\cdot)$ ensures that the threshold value is also monotone non-decreasing with increasing values of C . When $W(h) < C, \forall h$, we choose h to be ∞ , as done in Appendix G. This completes the proof of *indexability* for the decoupled problem. Observe that $C = W(h+1)$ is the minimum value of the activation charge that makes both actions equally desirable in state h . This gives us the expression for the Whittle index.

I. Proof of Theorem 9

Suppose that the base station knows that the i th process was at state $x_i(\tau) = x_0$ at time τ . Further, suppose that it received no additional updates regarding the i th process up to time-slot $\tau + \Delta$. Without loss of generality, we can set $\tau = 0$, since we can always offset the time-slots by a fixed constant.

Then, using the state evolution equation 13, we know that

$$\begin{aligned} x_i(1) &= G_i x_0 + w_i(0) \\ x_i(2) &= G_i^2 x_0 + G_i w_i(0) + w_i(1) \\ &\dots \\ x_i(\Delta) &= G_i^\Delta x_0 + \sum_{k=0}^{\Delta-1} G_i^{\Delta-k-1} w_i(k). \end{aligned} \quad (55)$$

The base station does not have access to the increments $w_i(0), \dots, w_i(\Delta-1)$. However, it knows that each of them is

i.i.d. and $\mathcal{N}(0, \Sigma_i)$. Thus, the maximum likelihood of the state at time Δ is given by

$$\hat{x}_i(\Delta) = \mathbb{E}[x_i(\Delta) | x_i(0) = x_0] = G_i^\Delta x_0. \quad (56)$$

Using this, we can now compute the difference between the actual state and the estimate at the base station

$$x_i(\Delta) - \hat{x}_i(\Delta) = \sum_{k=0}^{\Delta-1} G_i^{\Delta-k-1} w_i(k). \quad (57)$$

Observe that this is simply a sum of zero-mean independent multi-variate normal random variables. Thus, $x_i(\Delta) - \hat{x}_i(\Delta)$ is also a zero-mean multi-variate normal random variable.

Recall the following standard properties of multivariate normal random variables. If $X \sim \mathcal{N}(0, \Sigma)$ and $Y = GX$ is some linear transformation of X , then $Y \sim \mathcal{N}(0, G\Sigma G^T)$. Further, if $X_1 \sim \mathcal{N}(0, \Sigma_1)$ and $X_2 \sim \mathcal{N}(0, \Sigma_2)$ are independent, then $Z = X_1 + X_2$ is distributed as $\mathcal{N}(0, \Sigma_1 + \Sigma_2)$. Finally, if $X \sim \mathcal{N}(0, \Sigma)$, then $\mathbb{E}[X^T X] = \text{Tr}(\Sigma)$.

Putting the first two properties together, we observe that

$$x_i(\Delta) - \hat{x}_i(\Delta) \sim \mathcal{N}\left(0, \sum_{k=0}^{\Delta-1} G_i^k \Sigma_i (G_i^k)^T\right). \quad (58)$$

Using the last property, we get

$$\begin{aligned} e_i(\Delta) &= \mathbb{E}[(x_i(\Delta) - \hat{x}_i(\Delta))^T (x_i(\Delta) - \hat{x}_i(\Delta))] \\ &= \mathbb{E}[\|x_i(\Delta) - \hat{x}_i(\Delta)\|_2^2] \\ &= \text{Tr}\left(\sum_{k=0}^{\Delta-1} G_i^k \Sigma_i (G_i^k)^T\right) \\ &= \sum_{k=0}^{\Delta-1} \text{Tr}((G_i^k)^T \Sigma_i (G_i^k)) \\ &= \sum_{k=0}^{\Delta-1} \text{Tr}((G_i^k)^T (G_i^k) \Sigma_i) \triangleq f_i(\Delta). \end{aligned} \quad (59)$$

The last two equalities follow from the linearity of the trace operator and the fact that $\text{Tr}(AB) = \text{Tr}(BA)$. This completes the proof of Theorem 9.

We also want to show that $f_i(\Delta)$ increases monotonically in Δ . This is straightforward to show, since $G_i^k \Sigma_i (G_i^k)^T$ is a covariance matrix for any $k \in \mathbb{Z}^+$. This implies that it must be positive semi-definite, and in turn, must have a non-negative trace. Now, we consider the difference

$$f_i(\Delta+1) - f_i(\Delta) = \text{Tr}((G_i^\Delta) \Sigma_i (G_i^\Delta)^T) \geq 0. \quad (60)$$

The last inequality follows due to the non-negativity of trace for a positive semi-definite matrix. This shows that $f_i(\Delta+1) \geq f_i(\Delta)$, which allows us to conclude monotonicity of the AoI cost functions.

J. Proof of Theorem 10

As we discussed earlier, the estimate distribution at time $\tau + \Delta$, given the last observation at time τ is $x_i(\tau)$, is given by

$$\hat{x}_i(\tau + \Delta) = x_i(\tau) Q_i^\Delta.$$

Without loss of generality, we can assume that $x_i(\tau) = [1 \ 0]$, i.e. the chain at time τ is in state 0. This is because the chain is symmetric, so it does not matter which state we start from. Using this, we get the estimate distribution to be

$$\hat{x}_i(\tau + \Delta) = [[Q_i^\Delta]_{00} \ 1 - [Q_i^\Delta]_{00}]. \quad (61)$$

Further, at time $\tau + \Delta$ the actual state of the chain is 0 with probability $[Q_i^\Delta]_{00}$ and 1 with probability $1 - [Q_i^\Delta]_{00}$. Thus, the distribution of the actual state at time $\tau + \Delta$ is given by

$$x_i(\tau + \Delta) = \begin{cases} [1 \ 0], & \text{with probability } [Q_i^\Delta]_{00} \\ [0 \ 1], & \text{with probability } 1 - [Q_i^\Delta]_{00}. \end{cases} \quad (62)$$

Now, suppose that the distance between the actual and estimate distributions is measured using the Kullback-Leibler (KL) divergence. Then,

$$\begin{aligned} & \mathbb{E} \left[D_{KL}(x_i(\tau + \Delta) || \hat{x}_i(\tau + \Delta)) \right] \\ &= [Q_i^\Delta]_{00} D_{KL} \left([1 \ 0] || [[Q_i^\Delta]_{00} \ 1 - [Q_i^\Delta]_{00}] \right) \\ &+ (1 - [Q_i^\Delta]_{00}) D_{KL} \left([0 \ 1] || [[Q_i^\Delta]_{00} \ 1 - [Q_i^\Delta]_{00}] \right) \\ &= -[Q_i^\Delta]_{00} \log([Q_i^\Delta]_{00}) - (1 - [Q_i^\Delta]_{00}) \log(1 - [Q_i^\Delta]_{00}) \\ &= H([Q_i^\Delta]_{00}). \end{aligned} \quad (63)$$

Here $H(q) \triangleq -q \log(q) - (1 - q) \log(1 - q)$ is the binary entropy function.

Now, suppose that the distance between the actual and estimate distributions is measured using the total variation (TV) distance. Then,

$$\begin{aligned} & \mathbb{E} \left[D_{TV}(x_i(\tau + \Delta) || \hat{x}_i(\tau + \Delta)) \right] \\ &= [Q_i^\Delta]_{00} D_{TV} \left([1 \ 0] || [[Q_i^\Delta]_{00} \ 1 - [Q_i^\Delta]_{00}] \right) \\ &+ (1 - [Q_i^\Delta]_{00}) D_{TV} \left([0 \ 1] || [[Q_i^\Delta]_{00} \ 1 - [Q_i^\Delta]_{00}] \right) \\ &= [Q_i^\Delta]_{00} (1 - [Q_i^\Delta]_{00}) + (1 - [Q_i^\Delta]_{00}) [Q_i^\Delta]_{00} \\ &= 2[Q_i^\Delta]_{00} (1 - [Q_i^\Delta]_{00}) \\ &\triangleq g([Q_i^\Delta]_{00}). \end{aligned} \quad (64)$$

Here $g(x) = 2x(1 - x)$. This completes the proof of Theorem 10.

In addition, we also need to show that the two functions derived above are monotonically increasing. To do so, we will simplify our notation a bit. Let $\mu_0 = [Q_i^\Delta]_{00}$ by μ_0 and $\mu_1 = 1 - \mu_0 = [Q_i^\Delta]_{01}$. Further, let $\nu_0 = [Q_i^{\Delta+1}]_{00} = \mu_0(1 - q_i) + (1 - \mu_0)q_i$ and $\nu_1 = [Q_i^{\Delta+1}]_{01} = 1 - \nu_0 = \mu_1(1 - q_i) + (1 - \mu_1)q_i$. We will split the proof into two cases.

Case 1 (KL Divergence): Note that the function $x \log(x)$ is convex for all $x > 0$, since $\frac{d^2}{dx^2}(x \log(x)) = \frac{1}{x} > 0, \forall x > 0$. Using this fact and the definitions of ν_0 and ν_1 , we obtain the following inequalities:

$$(1 - q_i)\mu_0 \log(\mu_0) + q_i\mu_1 \log(\mu_1) \geq \nu_0 \log(\nu_0), \quad (65)$$

$$(1 - q_i)\mu_1 \log(\mu_1) + q_i\mu_0 \log(\mu_0) \geq \nu_1 \log(\nu_1), \quad (66)$$

Now, we look at the difference:

$$\begin{aligned} & H([Q_i^{\Delta+1}]_{00}) - H([Q_i^\Delta]_{00}) = \\ & \left((1 - q_i)\mu_0 \log(\mu_0) + q_i\mu_1 \log(\mu_1) - \nu_0 \log(\nu_0) \right) \\ & + \left((1 - q_i)\mu_1 \log(\mu_1) + q_i\mu_0 \log(\mu_0) - \nu_1 \log(\nu_1) \right) \\ & \geq 0. \end{aligned} \quad (67)$$

The inequality above follows by applying (65) and (66). This proves that the monitoring error grows monotonically with the AoI for KL divergence.

Case 2 (TV distance): Note that the function $2x(1 - x)$ is concave for all x , since $\frac{d^2}{dx^2}(2x(1 - x)) = -2 < 0, \forall x$. Using this fact and the definitions of ν_0 and ν_1 , we obtain the following inequalities:

$$(1 - q_i)2\mu_0(1 - \mu_0) + q_i2\mu_1(1 - \mu_1) \leq 2\nu_0(1 - \nu_0), \quad (68)$$

$$(1 - q_i)2\mu_1(1 - \mu_1) + q_i2\mu_0(1 - \mu_0) \leq 2\nu_1(1 - \nu_1), \quad (69)$$

Now, we look at the difference:

$$\begin{aligned} & g([Q_i^{\Delta+1}]_{00}) - g([Q_i^\Delta]_{00}) = \\ & \left(2\nu_0(1 - \nu_0) - (1 - q_i)2\mu_0(1 - \mu_0) - q_i2\mu_1(1 - \mu_1) \right) \\ & + \left(2\nu_1(1 - \nu_1) - (1 - q_i)2\mu_1(1 - \mu_1) - q_i2\mu_0(1 - \mu_0) \right) \\ & \geq 0. \end{aligned} \quad (70)$$

The inequality above follows by applying (68) and (69). This proves that the monitoring error grows monotonically with the AoI for TV distance as well.